

Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

5-1-1988

Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models

Donald W.K. Andrews

Follow this and additional works at: <https://elischolar.library.yale.edu/cowles-discussion-paper-series>



Part of the [Economics Commons](#)

Recommended Citation

Andrews, Donald W.K., "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models" (1988). *Cowles Foundation Discussion Papers*. 1117.
<https://elischolar.library.yale.edu/cowles-discussion-paper-series/1117>

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 874R

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than acknowledgment that a writer had access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

ASYMPTOTIC NORMALITY OF SERIES ESTIMATORS
FOR NONPARAMETRIC AND
SEMIPARAMETRIC REGRESSION MODELS

by

Donald W.K. Andrews

December, 1987
Revised: June, 1989

ASYMPTOTIC NORMALITY OF SERIES ESTIMATORS FOR NONPARAMETRIC AND SEMIPARAMETRIC REGRESSION MODELS

Donald W. K. Andrews¹

Cowles Foundation
Yale University

December, 1987

Revised: June, 1989

ABSTRACT

This paper establishes the asymptotic normality of series estimators for nonparametric regression models. Gallant's Fourier flexible form estimators, trigonometric series estimators, and polynomial series estimators are prime examples of the estimators covered by the results. The results apply to a wide variety of estimands in the regression model under consideration, including derivatives and integrals of the regression function. The errors in the model may be homoskedastic or heteroskedastic.

The paper also considers series estimators for additive interactive regression (AIR), semiparametric regression, and semiparametric index regression models and shows them to be consistent and asymptotically normal. All of the consistency and asymptotic normality results in the paper follow from one set of general results for series estimators.

1. INTRODUCTION

Recently, series estimators of different sorts and for different models have received considerable attention in the econometric and statistical literature. Gallant has introduced a class of *Fourier flexible form* (FFF) series estimators and has used them in a number of applications, see Gallant (1987) for references. Trigonometric series estimators of nonparametric regression models have been analyzed by numerous researchers, e.g., see Geman and Hwang (1982), Bergstrom (1985), Eastwood and Gallant (1987), Eastwood (1987, 1988), Severini and Wong (1987), Eubank (1988), and Cox (1988). Similarly, polynomial series estimators for various models have been considered, e.g., see Kendall and Stuart (1979, p. 380), Bergstrom (1985), Cox (1988), Fabian (1988), and Newey (1988a, b).

The purpose of this paper is to establish some asymptotic distribution theory for such series estimators in nonparametric regression models and several semiparametric models. Some asymptotic distribution theory is already available for these estimators. Most of the results concern the asymptotic distribution of trigonometric series estimators of the regression function in a particular nonparametric regression model, see Eastwood (1987, 1988), Eastwood and Gallant (1987), and Eubank (1988). The model in question has a single regressor that takes on equi-spaced values in $[0, 2\pi]$, a regression function that is periodic on $[0, 2\pi]$, and errors that are independent and identically distributed (iid). Eastwood's (1987, 1988) and Eastwood and Gallant's (1987) results also cover the estimation of derivatives of the regression function in this model.

Other asymptotic distributional results in the literature for series estimators include those of Chamberlain (1986), who considers a semiparametric regression model, Souza (1987), who considers a nonparametric multivariate regression model, and Newey (1988b), who considers a sample selectivity model. The other papers referred to above do not give distributional results for series estimators.

In this paper, we establish the asymptotic normality of series estimators in various nonparametric and semiparametric regression models. Fourier flexible form (FFF), trigonometric, and polynomial series estimators are covered by the results. Multiple regressors are allowed, the values of the regressors are not restricted, and the errors may be independent non-identically distributed (inid). Both fixed and adaptive truncation sequences are considered. Results for the latter follow the approach of Eastwood and Gallant (1987), but are more general.

The estimands considered here are diverse. They include the regression function itself, derivatives of the regression function, integrated values of the regression function, average values of the derivatives of the regression function over the points in the sample, and smooth functions of the aforementioned estimands. Consistent covariance matrix estimators are provided.

The models considered here include the standard nonparametric regression model, additive interactive regression (AIR) models (also known as interaction spline models, see Barry (1983, 1986) and Wahba (1986)), the semiparametric regression model (e.g., see Robinson (1988) and Chamberlain (1986, 1987)), and semiparametric index regression models (e.g., see Ruud (1986) and Stoker (1986)).

One of the attributes of this paper is that the asymptotic normality results for the different estimands, estimators, and models considered are all obtained from a single set of results for series estimators.

The remainder of this paper is organized as follows: Section 2 deals with nonparametric regression models and treats series estimators that are based on non-random regressors and non-random truncation sequences. Section 3 presents results that extend the asymptotic normality results of Section 2 to models with random regressors. Section 4 extends the results of Section 2 to series estimators that are based on certain random, data-dependent truncation sequences. Section 5 considers series estimators of AIR models, semiparametric regression models, and semiparametric index regression models. Section 6

considers the estimation of covariance matrices of series estimators for each of the above models when the errors are homoskedastic or heteroskedastic. An Appendix contains proofs of results stated in the text.

2. ASYMPTOTIC NORMALITY IN NONPARAMETRIC REGRESSION MODELS

This section is divided into three subsections. The first defines the model, the estimands, and the estimators of interest, the second presents the assumptions used and the asymptotic normality results, and the third discusses the assumptions.

2.1. Definition of the Model, Estimands, and Estimators

The model considered here is given by

$$(2.1) \quad Y_i = g(x_i) + U_i, \quad i = 1, \dots, n,$$

where $Y_i, U_i \in \mathbb{R}$, $x_i \in \mathcal{X} \subset \mathbb{R}^d$, $g(\cdot) \in \mathcal{G}$, and \mathcal{G} is a class of continuous functions from \mathcal{X} to \mathbb{R} .

The objective is to estimate various functions of $g(\cdot)$, such as $g(x)$ and derivatives of $g(x)$ for arbitrary $x \in \mathcal{X}$. The approach taken is to approximate $g(\cdot)$ by a series $\sum_{s=1}^{\kappa_n} z_s(\cdot) \theta_s$, where $\{z_s(\cdot) : s = 1, 2, \dots\}$ is a prespecified family of functions from \mathcal{X} to \mathbb{R} , $\theta_{\kappa_n} = (\theta_1, \dots, \theta_{\kappa_n})'$ is an unknown parameter vector, and κ_n is the number of summands in the series when the sample size is n . In this section κ_n is taken to be non-random. In Section 4, random data-dependent values of κ_n are considered. The results given below are stated so that they apply to any family $\{z_s(\cdot)\}$ that satisfies certain properties. Families of particular interest include the Fourier flexible form (FFF), trigonometric, and polynomial families.

The FFF family contains linear, quadratic, and trigonometric functions of the regressor vector $x = (x_1, \dots, x_d)'$. Thus, $z_s(x)$ for $s \geq 1$ is of the form:

$$(2.2) \quad \begin{aligned} &1, x_r \text{ for } r = 1, \dots, d, x_r x_\ell \text{ for } r, \ell = 1, \dots, d, \text{ or} \\ &\cos(b'x) \text{ or } \sin(b'x) \text{ for some vector } b \in R^d \text{ with integer elements.} \end{aligned}$$

A precise enumeration and ordering of the functions in the FFF family is given in Gallant (1981, p. 219). The trigonometric family is the same as the FFF family except the linear and quadratic functions of the elements of x are omitted. The polynomial family contains polynomial functions of the elements of the regressor vector x .

The FFF and trigonometric families are designed to approximate functions over the domain $[0, 2\pi]^d$. Hence, one has to shift and rescale the regressors so that each element of the regressor vector lies in $[0, 2\pi]$ before carrying out the estimation procedure.

The parameter vector θ_{κ_n} is estimated by least squares (LS). To define the LS estimator, we introduce some notation: Throughout this paper, k denotes some fixed element of I_+ , the set of positive integers, and κ denotes a function from I_+ to I_+ . κ is called a *truncation sequence* because $\kappa(n) (= \kappa_n)$ denotes the number of terms to be included in the series expansion when the sample size is n . Let

$$(2.3) \quad \begin{aligned} Z_k(\cdot) &= (z_1(\cdot), \dots, z_k(\cdot))' \in \mathcal{G}^k, \\ Z_{nk} &= (Z_k(x_1), \dots, Z_k(x_n))' \in R^{n \times k}, \end{aligned}$$

$$Y = (Y_1, \dots, Y_n)' \in R^n, \text{ and } U = (U_1, \dots, U_n)' \in R^n.$$

For a truncation sequence κ , let $Z_{n\kappa}$ abbreviate $Z_{n\kappa_n}$. When the sample size n and the truncation sequence κ are evident from the context, let Z abbreviate $Z_{n\kappa}$.

The LS estimator of θ_{κ_n} is

$$(2.4) \quad \hat{\theta}_{n\kappa} = (Z'Z)^+ Z'Y,$$

where $(\cdot)^+$ denotes the Moore–Penrose inverse. The corresponding *series estimator* \hat{g} of g is

$$(2.5) \quad \hat{g}(\cdot) = Z_{\kappa_n}(\cdot)' \hat{\theta}_{n\kappa}.$$

Next we describe the functions of g that we wish to estimate. Let $\Gamma_n(g)$ denote the estimand, where $\Gamma_n(\cdot)$ is a function from \mathcal{G} to \mathbb{R}^V . Let $\lambda_j = (\lambda_{j1}, \dots, \lambda_{jd})'$ be a vector of non-negative integers. Define

$$(2.6) \quad |\lambda_j| = \sum_{r=1}^d \lambda_{jr} \text{ and } D^{\lambda_j} g(x_c) = \frac{\partial^{|\lambda_j|}}{\partial x_{c1}^{\lambda_{j1}} \dots \partial x_{cd}^{\lambda_{jd}}} g(x_c),$$

where $x_c = (x_{c1}, \dots, x_{cd})' \in \mathcal{X}$. The function $\Gamma_n(\cdot)$ is defined so as to include the following examples:

1. $\Gamma_n(g) = g(x)$ for some $x \in \mathcal{X}$,
2. $\Gamma_n(g) = (g(x_1), \dots, g(x_v))'$ for some $(x'_1, \dots, x'_v)' \in \mathcal{X}^V$,
3. $\Gamma_n(g) = \int_{\mathcal{X}} g(x) d\eta(x)$ for some probability distribution $\eta(\cdot)$ on \mathcal{X} ,
- (2.7) 4. $\Gamma_n(g) = (D^{\lambda_1} g(x_1), \dots, D^{\lambda_v} g(x_v))'$ for some $(x'_1, \dots, x'_v)' \in \mathcal{X}^V$,
5. $\Gamma_n(g) = \frac{1}{n} \sum_{i=1}^n (D^{\lambda_1} g(x_i), \dots, D^{\lambda_v} g(x_i))'$ for the observed regressor vectors (x_1, \dots, x_n) , and
6. $\Gamma_n(g) = \left[\int_{\mathcal{X}} D^{\lambda_1} g(x) d\eta_1(x), \dots, \int_{\mathcal{X}} D^{\lambda_v} g(x) d\eta_v(x) \right]'$ for some probability distributions η_1, \dots, η_v on \mathcal{X} (which could depend on n).

See Stoker (1986, 1987), Powell, Stock, and Stoker (1989), Härdle and Stoker (1987), and Section 5.3 below for index model examples that motivate interest in Example 5 and for alternative estimation methods for it.

The estimator of $\Gamma_n(g)$ that we consider is $\Gamma_n(\hat{g})$. Since the functional $\Gamma_n(\cdot)$ is assumed below to be linear, we have

$$\begin{aligned}
(2.8) \quad & \Gamma_n(\hat{g}) = \Gamma_n(Z_{\kappa_n}(\cdot))' \hat{\theta}_{n\kappa} = \gamma'_{n\kappa} \hat{\theta}_{n\kappa}, \text{ where} \\
& \gamma_n(s) = \Gamma_n(z_s(\cdot)) \in \mathbb{R}^v \quad \forall s \geq 1, \quad \gamma_{n\kappa} = (\gamma_n(1), \dots, \gamma_n(k))' \in \mathbb{R}^{k \times v}, \\
& \text{and } \gamma_{n\kappa} \text{ abbreviates } \gamma_{n\kappa_n}. \text{ For example, in Example 1 of (2.7)} \\
& \gamma_{n\kappa} = (z_1(x), \dots, z_{\kappa_n}(x))', \text{ in Example 2 the } j\text{-th column of } \gamma_{n\kappa} \text{ is} \\
& (z_1(x_j), \dots, z_{\kappa_n}(x_j))' \text{ for } j = 1, \dots, v, \text{ in Example 3} \\
& \gamma_{n\kappa} = \left[\int_{\mathcal{X}} z_1(x) d\eta(x), \dots, \int_{\mathcal{X}} z_{\kappa_n}(x) d\eta(x) \right]', \text{ and in Example 4 the } j\text{-th column of } \gamma_{n\kappa} \\
& \text{is } (D^{\lambda_j} z_1(x_j), \dots, D^{\lambda_j} z_{\kappa_n}(x_j))' \text{ for } j = 1, \dots, v.
\end{aligned}$$

2.2. Assumptions and Asymptotic Normality Results

First we introduce some notation and definitions used in the assumptions and Theorem 1 below. Let σ_i^2 denote the variance of U_i for $i = 1, \dots, n$. Let $\Omega = \text{diag}\{\sigma_1^2, \dots, \sigma_n^2\}$ ($= EUU'$). Define

$$\begin{aligned}
(2.9) \quad & V_n = \text{Var}(\Gamma_n(\hat{g})) = \gamma'_{n\kappa} (Z'Z)^+ Z' \Omega Z (Z'Z)^+ \gamma_{n\kappa} \text{ and} \\
& A_n = V_n^{-1/2}.
\end{aligned}$$

For any $g_1 \in \mathcal{G}$, let $\|g_1\|_{q, \omega, \mathcal{X}}$ denote the supremum Sobolev norm of derivative order q for some nonnegative integer q . That is,

$$(2.10) \quad \|g_1\|_{q, \omega, \mathcal{X}} = \sum_{\lambda: |\lambda| \leq q} \sup_{x \in \mathcal{X}} |D^\lambda g_1(x)|,$$

where $\lambda \in I_+^d$, $|\lambda|$, and $D^\lambda g_1(x)$ are as in (2.6). If partial derivatives of $g_1(x)$ do not exist up to order q , then $\|g_1\|_{q, \omega, \mathcal{X}} = \infty$.

Let $PZ_{n\kappa}(i)$ denote the i -th diagonal element of the $n \times n$ projection matrix $Z_{n\kappa}(Z'_{n\kappa} Z_{n\kappa})^+ Z'_{n\kappa}$. Let $PZ_{n\kappa}(i)$ abbreviate $PZ_{n\kappa_n}(i)$ for any truncation sequence κ .

Let $\zeta(\cdot): I_+ \rightarrow \mathbb{R} \cup \{\infty\}$ be a non-decreasing function for which

$$(2.11) \quad \zeta(s) \geq \sup_{x \in \mathcal{X}} |z_s(x)| \quad \forall s \geq 1.$$

For example, one could take $\zeta(s) = \max_{m \leq s} \sup_{x \in \mathcal{X}} |z_m(x)|$. For certain results below, $\zeta(s)$ is assumed to be finite for all $s \geq 1$. In fact, for each of the families of series functions referred to above, $\zeta(\cdot)$ can be taken to be a finite constant provided \mathcal{X} is bounded.

Let $\kappa^{-1} : I_+ \rightarrow I_+$ denote the *inverse function* of the truncation sequence κ . It is defined such that $\kappa^{-1}(\kappa_n) \geq n \quad \forall n \geq 1$. In particular, let

$$(2.12) \quad \kappa^{-1}(k) = \min\{n \in I_+ : \kappa_m > k \quad \forall m > n\} \quad \forall k \geq 1.$$

If $\kappa_n \rightarrow \infty$ as $n \rightarrow \infty$, then κ^{-1} is well-defined on I_+ . For example, if $\kappa_n = [n^r] \quad \forall n \geq 1$ for some $r \in (0,1)$, where $[\cdot]$ denotes the integer part of \cdot , then $\kappa^{-1}(k) = [k^{1/r}] + \delta_k \quad \forall k \geq 1$, where $\delta_k = 0$ if $k^{1/r} \in I_+$ and $\delta_k = 1$ otherwise.

For any finite dimensional vector or matrix b , let $\|b\|$ denote the Euclidean norm of b . For any square matrix B , let $\lambda_{\min}(B)$ and $\lambda_{\max}(B)$ denote the minimum and maximum eigenvalues of B respectively.

We now state the assumptions—each of which is discussed below in Section 2.3. The first three assumptions, A–C, are sufficient for asymptotic normality of $A_n(\Gamma_n(\hat{g}) - E\Gamma_n(\hat{g}))$; the first four, A–D, are sufficient for asymptotic normality of $A_n(\Gamma_n(\hat{g}) - \Gamma_n(g))$. The fifth, E, is necessary and sufficient for the variance of $\Gamma_n(\hat{g})$ to go to zero as $n \rightarrow \infty$. The final two assumptions, F and G, each provide easily verifiable sufficient conditions for Assumptions C and E.

ASSUMPTION A: $\{U_i : i \geq 1\}$ are independent rv's with (i) $EU_i = 0 \quad \forall i$,
(ii) $0 < \inf_{i \geq 1} \sigma_i^2 \leq \sup_{i \geq 1} \sigma_i^2 < \infty$, and (iii) $\sup_{i \geq 1} EU_i^2 1(|U_i| > c) \rightarrow 0$ as $c \rightarrow \infty$.
 $\{x_i : i \geq 1\}$ are non-stochastic regressor vectors in $\mathcal{X} \subset \mathbb{R}^d$.

ASSUMPTION B: (i) $\Gamma_n(\cdot)$ is a linear functional.

(ii) For some $C_1 < \infty$, $\epsilon > 0$, and integer $q \geq 0$, $\Gamma_n(\cdot)$ satisfies: $\forall g_1 \in \mathcal{G}$ with $\|g_1\|_{q,\omega,\lambda} < \epsilon$, $\|\Gamma_n(g_1)\| \leq C_1 \|g_1\|_{q,\omega,\lambda}$ for all n large.

(iii) $\lim_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{nk} \gamma_{nk}) > 0$ for some $k \geq 1$.

ASSUMPTION C: (i) $\kappa_n \rightarrow \infty$ as $n \rightarrow \infty$. (ii) $\max_{i \leq n} PZ_{nk}(i) \rightarrow 0$ as $n \rightarrow \infty$. (iii) Either

$Z_{n\kappa}$ is full rank κ_n for n large or $\Gamma_n(g) = g(x_j)$ for some regressor x_j that is observed for n large and $z_s(x_j) \neq 0$ for some $s \geq 1$.

ASSUMPTION D: The regression function g satisfies: $\forall k \geq 1$

$\exists \theta_k = (\theta_{k1}, \dots, \theta_{kk})' \in \mathbb{R}^k$ such that

$$(2.13) \quad \left[\kappa^{-1}(k)k \right]^{1/2} \left\| \sum_{s=1}^k z_s(\cdot) \theta_{ks} - g(\cdot) \right\|_{q,\omega,\lambda} \rightarrow 0 \text{ as } k \rightarrow \infty.$$

If $\lim_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{n\kappa} (Z'Z/n)^+ \gamma_{n\kappa}) > 0$, then $\left[\kappa^{-1}(k)k \right]^{1/2} \zeta(k)$ can be replaced by $\kappa^{-1}(k)^{1/2}$ in (2.13).³

ASSUMPTION E: $\gamma'_{n\kappa} (Z'Z)^+ \gamma_{n\kappa} \rightarrow 0$ as $n \rightarrow \infty$.

We now introduce two upper bound functions κ_b and κ_b^* such that (i) if $\kappa_n \leq \kappa_{bn}$ for all n large, $\Gamma_n(g) = g(x_j)$ for some observed regressor x_j , and Assumption F (below) holds, then Assumptions C and E hold and (ii) if $\kappa_n \leq \kappa_{bn}^*$ for all n large and Assumption G (below) holds, then Assumptions C and E hold. Suppose $b(\cdot)$ is a positive real function that satisfies $b(x) \rightarrow 0$ as $x \rightarrow \infty$. Define the function $\kappa_b : I_+ \rightarrow I_+$ as follows:

$$(2.14) \quad \kappa_{bn} = \max\{k \in I_+ : \max_{i \leq n} PZ_{nk}(i) \leq b(k)\}$$

and $\kappa_{bn} = 1$ if the set in (2.14) is empty. Under Assumption F below, for each fixed k the set in (2.14) is non-empty for n large. In consequence, $\kappa_{bn} \rightarrow \infty$ as $n \rightarrow \infty$.

Define the function $\kappa_b^* : I_+ \rightarrow I_+$ as follows:

$$(2.15) \quad \kappa_{bn}^* = \max\{k \in I_+ : \max_{i \leq n} PZ_{nk}(i) \leq b(k), \max_{m \leq k} \lambda_{\max}(\gamma'_{nm}(Z'_{nm}Z_{nm})^+ \gamma_{nm}) \leq b(k), \\ \text{and } Z_{nk} \text{ is full rank } k\}$$

and $\kappa_{bn}^* = 1$ if the set in (2.15) is empty. When $\Gamma_n(g) = g(x_j)$ for some regressor x_j that is observed for n large, the full rank condition in (2.15) is deleted. Under Assumption G below, $\kappa_{bn}^* \rightarrow \infty$ as $n \rightarrow \infty$ (for the same reason as for κ_{bn}).

ASSUMPTION F: (i) $\Gamma_n(g) = g(x_j)$ for some regressor x_j that is observed for n large and $z_s(x_j) \neq 0$ for some $s \geq 1$.

(ii) $\kappa_n \leq \kappa_{bn}$ for all n large and $\kappa_n \rightarrow \infty$ as $n \rightarrow \infty$.

(iii) $\frac{1}{n} \sum_{i=1}^n Z_k(x_i)Z_k(x_i)' \rightarrow \int_{\mathcal{X}} Z_k(x)Z_k(x)' dF(x)$ as $n \rightarrow \infty \quad \forall k \geq 1$ for some distribution function $F(\cdot)$ on \mathcal{X} .

(iv) If $\int_{\mathcal{X}} (c'Z_k(x))^2 dF(x) = 0$ for some $k \geq 1$ and $c \in \mathbb{R}^k$, then $c'Z_k(x_i) = 0 \quad \forall i \geq 1$.

(v) $\sup_{x \in \mathcal{X}} |z_s(x)| < \infty \quad \forall s \geq 1$.

ASSUMPTION G: (i) $\kappa_n \leq \kappa_{bn}^*$ for all n large and $\kappa_n \rightarrow \infty$ as $n \rightarrow \infty$.

(ii) $\lambda_{\min}(Z'_{nk}Z_{nk}) \rightarrow \infty$ as $n \rightarrow \infty \quad \forall k \geq 1$.

(iii) $\sup_{x \in \mathcal{X}} |z_s(x)| < \infty \quad \forall s \geq 1$.

(iv) $\overline{\lim}_{n \rightarrow \infty} \lambda_{\max}(\gamma'_{nk}\gamma_{nk}) < \infty \quad \forall k \geq 1$.

THEOREM 1: (a) Under Assumptions A–C,

$$A_n(\Gamma_n(\hat{g}) - E\Gamma_n(\hat{g})) \xrightarrow{d} N(0, I_v) \text{ as } n \rightarrow \infty.$$

(b) Under Assumptions A–D,

$$A_n(\Gamma_n(\hat{g}) - \Gamma_n(g)) \xrightarrow{d} N(0, I_v) \text{ as } n \rightarrow \infty.$$

(c) Under Assumptions A–D,

$$\Gamma_n(\hat{g}) - \Gamma_n(g) \xrightarrow{P} 0 \text{ as } n \rightarrow \infty$$

if and only if Assumption E holds. The same is true with " \xrightarrow{P} " replaced by " $\xrightarrow{L^2}$ ".

(d) Assumption F implies Assumptions C and E.

(e) Assumption G implies Assumptions C and E.

COMMENTS: 1. The conditions of Assumptions A, B, F, and G are quite mild. In consequence, Theorem 1(a), (d), and (e) show that under mild conditions it is always possible to obtain truncation sequences κ such that $\Gamma_n(\hat{g})$ is asymptotically normal with centering at its expectation for a wide variety of estimands $\Gamma_n(g)$. Furthermore, Assumption D is a smoothness condition on the regression function g (see Section 2.3 below) that is always satisfied for an infinite class of functions in \mathcal{G} . Thus, parts (b), (c), (d), and (e) of Theorem 1 can be interpreted as saying: Under mild conditions there exist truncation sequences κ such that $\Gamma_n(\hat{g})$ is consistent and asymptotically normal with centering at the estimand $\Gamma_n(g)$ for a wide variety of estimands $\Gamma_n(g)$, provided g is sufficiently smooth. An important practical consequence of Theorem 1(d) and (e) is that such truncation sequences κ are easily obtainable in practice. Given a function $b(\cdot)$ as above, the upper bounds κ_{bn} and κ_{bn}^* on κ_n are observable and readily computable, since they only depend on the regressor values, the prespecified series functions, and the known function $\Gamma_n(\cdot)$ that defines the estimand.

2. A drawback of Theorem 1(c), on the other hand, is that the smoothness Assumption D can be difficult to interpret. If the rate of increase of κ_n or of a lower bound on κ_n is not known explicitly, then one cannot be explicit regarding the necessary rate of decay of the series coefficients of g for Assumption D to hold. For example, if one imposes Assumptions C and E by requiring $\kappa_n \leq \kappa_{bn}$ or $\kappa_n \leq \kappa_{bn}^*$ and the rate of increase of κ_{bn} or κ_{bn}^* is not explicit (but rather, is only known as a function of $b(\cdot)$, $PZ_{nk}(i)$, and $\gamma'_{nk}(Z'Z)^+ \gamma_{nk}$ for $k \geq 1$), then a rate of increase of κ_n or of a lower bound on κ_n

cannot be made explicit. Section 3 below is devoted to obtaining sufficient conditions under which more explicit growth rates of κ_n are obtained.

3. Although Assumption A takes the regressors $\{x_i\}$ to be non-random, Theorem 1 also holds for random regressors $\{X_i\}$ if the assumptions of the Theorem hold conditional on $\{X_i\} = \{x_i\}$ with probability one.⁴ In this case, σ_j^2 denotes $\text{Var}(U_j | \{X_i\} = \{x_i\})$ and all expectations that appear in the assumptions are conditional expectations given $\{X_i\} = \{x_i\}$. For example, if $\Gamma_n(g) = g(x_j)$ for some observed x_j , then Assumption B necessarily holds and Assumption F holds with probability one if the regressors $\{X_i\}$ are bounded iid rv's and FFF, trigonometric, or polynomial series functions are used (see Section 2.3 below). No other assumptions on the regressors are needed for the results of Theorem 1. For more general estimands, Assumption G holds with probability one under the same conditions plus $EZ_k(X_i)Z_k(X_i)'$ is nonsingular for all $k \geq 1$ (see Section 2.3).

4. By applying the delta method, Theorem 1 yields the asymptotic normality of series estimators of nonlinear functions $h(\cdot) : R^v \rightarrow R^w$ of $\{\Gamma_n(g)\}$. Consider the estimators $\{h(\Gamma_n(\hat{g}))\}$. Suppose $\frac{\partial}{\partial \xi} h(\cdot)$ exists, is full rank $w (\leq v)$, and is uniformly continuous at the points $\{\Gamma_n(g) : n \geq 1\}$. Let $B_n = \left[\frac{\partial}{\partial \xi} h(\Gamma_n(g)) V_n \left[\frac{\partial}{\partial \xi} h(\Gamma_n(g)) \right]' \right]^{-1/2}$. Then, under Assumptions A–E, we have $h(\Gamma_n(\hat{g})) - h(\Gamma_n(g)) \xrightarrow{P} 0$ as $n \rightarrow \infty$ and

$$(2.16) \quad B_n \left[h(\Gamma_n(\hat{g})) - h(\Gamma_n(g)) \right] \xrightarrow{d} N(0, I_w) \text{ as } n \rightarrow \infty.$$

This is proved using the same argument as is used to prove the delta method (e.g., see Bishop, Fienberg, and Holland (1975, pp. 486–497)).

5. The results of Theorem 1 and (2.16) above (coupled with the consistent covariance matrix estimators given in Section 6) allow one to construct Wald tests of a wide variety of nonlinear restrictions on the estimands under consideration. For example, one can test a number of different *derivative constraints* discussed by Stoker (1987).

6. The closer κ_n is to κ_{bn} or κ_{bn}^* of Assumptions F and G, the weaker is the smoothness assumption on g (i.e., Assumption D) that is needed for asymptotic normality

of $\Gamma_n(\hat{g}) - \Gamma_n(g)$. This does not mean, however, that one should necessarily choose κ_n close to or equal to κ_{bn} or κ_{bn}^* . Such a choice may yield a slow rate of convergence of $\Gamma_n(\hat{g}) - \Gamma_n(g)$ to 0.

7. If the estimand $\Gamma_n(g)$ is the regression function evaluated at a point x_j that is observed for some n large, i.e., $\Gamma_n(g) = g(x_j)$, then the assumptions above are coordinate free. That is, they only concern the linear space spanned by the columns of Z for n large and the linear space spanned by the functions $\{z_s(\cdot) : s = 1, \dots, k\}$ for k large. They do not concern the particular vectors and functions used to describe these spaces. Also, for this estimand, Assumption E is implied by Assumption C(ii) and the minimum eigenvalue condition in Assumption D holds provided $\{z_s(\cdot)\}$ contains a constant function (since $e_j' Z(Z'Z)^+ Z' e_j \geq e_j' \frac{1}{n} (1'1)^{-1} 1' e_j \geq 1/n$).

8. If the errors $\{U_i\}$ are normally distributed, then the exact distribution of $\Gamma_n(\hat{g})$ is multivariate normal with mean $E\Gamma_n(\hat{g})$ and covariance matrix A_n^{-2} . Thus, the normal approximation suggested by Theorem 1 will be quite accurate in many cases with respect to the variance and shape of the density of $\Gamma_n(\hat{g})$.

2.3. Discussion of the Assumptions

Here we discuss the strength and verifiability of Assumptions A–G. Assumption A allows for iid errors with finite variances. For cross-sectional situations, these assumptions are quite weak. Among others, they allow for applications to limited dependent variable regression models in which the errors are naturally heteroskedastic. As noted in Comment 3 following Theorem 1, random regressors can be handled by conditioning on the regressors. (This precludes, however, dynamic regression models.)

For time series situations, the independence assumption on the errors could be replaced by an assumption that allows for asymptotically weak temporal dependence. The conditions required in this case would be more complicated and the estimation of covariance matrices of the estimators would be more difficult. The proof of the results, however,

could be extended by using an appropriate central limit theorem.

The linearity of $\Gamma_n(\cdot)$ in Assumption B(i) is basic to the approach taken here. All of the examples in (2.7) satisfy this assumption. In addition, by using the delta method, countless nonlinear differentiable functions $h(\cdot) : R^V \rightarrow R^W$ of the estimands $\{\Gamma_n(g)\}$ can be shown to be asymptotically normal, see Comment 4 following Theorem 1.

Assumption B(ii) requires the linear functions $\Gamma_n(\cdot)$ to satisfy a Lipschitz condition at 0 uniformly over n and $\Gamma_n(0) = 0$. This assumption is usually easy to verify. For example, all of the examples of (2.7) satisfy it trivially. The integer q is equal to $0, 0, 0, |\lambda|, \max_{j \leq v} |\lambda_j|$, and $\max_{j \leq v} |\lambda_j|$ in Examples 1–6 of (2.7) respectively.

Assumption B(iii) implies that Theorem 1 only applies to vectors of estimands whose asymptotic distribution is nonsingular. If the estimand is vector-valued and Assumption B(iii) does not hold, then one or more elements of the estimand $\Gamma_n(g)$ need to be dropped before applying Theorem 1. If the estimand is a scalar, then B(iii) usually is trivial to verify. For example, if $\Gamma_n(\cdot)$ does not depend on n , then B(iii) holds if $\Gamma_n(z_s(\cdot)) \neq 0$ for some $s \geq 1$.

Assumptions C and E are discussed only briefly here, since either of Assumptions F and G is sufficient for both to hold and Assumption H of Section 3 below is sufficient for both to hold with probability one when the regressors are random. For convenience, Assumption D is discussed following the discussion of Assumptions C, E, F, and G.

Assumption C(i) is quite natural. Either g can be represented by a finite expansion in the series functions or the bias of $\Gamma_n(\hat{g})$ converges to zero as $n \rightarrow \infty$ only if $\kappa_n \rightarrow \infty$. The former case is covered by standard linear regression results and the latter case requires Assumption C(i).

Assumption C(ii) restricts the growth rate of κ_n . More precisely, it restricts the growth rate of the number of linearly independent columns of Z , which equals κ_n if $Z'Z$ is nonsingular. If $Z'Z$ is nonsingular, Assumption C(ii) implies that $\kappa_n/n \rightarrow 0$ as

$n \rightarrow \infty$, since $\max_{i \leq n} PZ_{n\kappa}(i) \geq \text{tr}(Z(Z'Z)^{-1}Z')/n = \kappa_n/n$. Note that Assumption C(ii) is coordinate free, since $\max_{i \leq n} PZ_{n\kappa}(i)$ depends on Z only through the projection matrix $Z(Z'Z)^+Z'$. Assumption C(ii) cannot be relaxed appreciably since it is a necessary condition for L^2 -consistency and asymptotic normality of certain sequences of estimators.

A sufficient condition for Assumption C(ii) is

$$(2.17) \quad \lambda_{\min}(Z'Z/(\kappa_n \zeta^2(\kappa_n))) \rightarrow \infty \text{ as } n \rightarrow \infty.$$

This condition is satisfied in the model and estimation procedure analyzed by Eastwood and Gallant (1987) provided $\kappa_n/n \rightarrow \infty$ as $n \rightarrow \infty$, since $\sup_{s \geq 1} |\zeta(s)| < \infty$ and $Z'Z = \text{diag}(n, n/2, \dots, n/2)$. Also, condition (2.17) holds with the choice of κ_n given in Theorem 2 of Section 3 below. Alternative (and more primitive) sufficient conditions for Assumption C(ii) are given in Assumptions F and G.

Assumption C(iii) guarantees that the LS estimator of $\Gamma_n(g)$ is unique. If $\Gamma_n(g) = g(x_j)$ for some observed x_j , then $\Gamma_n(\hat{g}) = \gamma'_{n\kappa} \hat{\theta}_{n\kappa} = e'_j Z(Z'Z)^+Z'Y$ and the value of $\Gamma_n(\hat{g})$ does not depend on the arbitrary choice of g -inverse $(\cdot)^+$. For other estimands, however, $\Gamma_n(\hat{g}) = \gamma'_{n\kappa} (Z'Z)^+Z'Y$ generally does depend on the choice of g -inverse unless Z is of full column rank.

Like Assumption C(ii), Assumption E restricts the growth rate of κ_n . In some cases (see Comment 7 following Theorem 1), Assumption E is implied by Assumption C(ii). In other cases, κ_n must be chosen with both C(ii) and E in mind.

Next, we discuss Assumptions F and G. Assumption F applies only to estimands of the form $\Gamma_n(g) = g(x_j)$, whereas Assumption G applies more generally. Truncation sequences κ that satisfy Assumptions F(ii) and G(i) always exist, since the upper bounds κ_{bn} and κ_{bn}^* on κ_n go to infinity as $n \rightarrow \infty$. These assumptions still may be restrictive, however, since they may impose a slow growth rate on κ_n . The slower the growth rate of κ_n , the smaller is the class of functions for which Assumption D holds.

Assumption F(iii) and (iv) hold with probability one if $\{X_i\}$ is an iid or stationary ergodic sequence (using the strong law of large numbers, the boundedness assumption of $F(v)$, and the fact that a rv with second moment zero equals zero with probability one). Furthermore, Assumption F(v) holds for most series functions defined on bounded sets. For example, it holds for FFF, trigonometric, and polynomial functions defined on bounded sets. More generally, it holds whenever \mathcal{X} is compact and $z_s(x)$ is continuous in x for all $s \geq 1$. (Note that Assumption F(v) does *not* require $z_s(x)$ to be bounded uniformly over x and s .) In sum, Assumption F is quite weak and easily verifiable. It holds for typical realizations of iid regressor sequences for any of the common choices of series functions.

Next, consider Assumption G. Assumption G(ii) is the well known necessary and sufficient condition for (weakly) consistent estimation of the parameters in a linear regression model with a fixed number of regressors k , regressor matrix Z_{nk} , and iid mean zero square integrable errors (see Drygas (1976, Thm. 3.1(a))) for each $k \geq 1$. It is implied by the standard assumption for linear regression models that $Z'_{nk}Z_{nk}/n$ converges to a positive definite matrix. If this condition is violated, then redundant series functions can be eliminated to make it hold. If the regressors $\{X_i\}$ are iid (or stationary and ergodic), then Assumption G(ii) holds with probability one if $EZ_k(X_i)Z_k(X_i)'$ is non-singular for all $k \geq 1$ (by the strong law of large numbers and Assumption G(iii)). Again, if the latter condition is violated, then series functions that are redundant in the limit can be eliminated. Knowledge of a lower bound on the density of X_i (with respect to some measure) can be used to determine which functions are redundant.

Assumption G(iii) is the same as Assumption F(v), which is discussed above. Assumption G(iv) is automatically satisfied in the common case where $\Gamma_n(\cdot)$ does not depend on n (see Examples 1–4 and 6 of (2.7)). Even if $\Gamma_n(\cdot)$ depends on n , Assumption G(iv) is usually satisfied. In Example 5 of (2.7), it holds if $\sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n (D^{\lambda_j} z_s(x_i))^2 < \infty$

$\forall s \geq 1$. The latter holds for FFF, trigonometric, and polynomial functions defined on bounded sets.

In sum, Assumption G is fairly weak and easily verified (when it holds). It is stronger than Assumption F in that redundant series functions need to be eliminated for it to be satisfied. For estimands not of the form $\Gamma_n(g) = g(x_j)$, redundant series functions generally cause non-uniqueness of the LS estimator $\Gamma_n(\hat{g})$ (see the discussion above of Assumption C(iii)). For such estimands, the elimination of redundant series functions seems difficult to avoid.

A final comment on Assumptions F and G: Although these conditions are reasonably weak, are not difficult to verify, and lead to easily implementable restrictions on κ_n that guarantee that Assumptions C and E hold, they do suffer from the drawback noted in Comment 2 following Theorem 1. Section 3 below gives some alternative sufficient conditions for Assumptions C and E, which are more restrictive than Assumptions F and G, but which lead to explicit expressions for $\kappa^{-1}(\cdot)$.

We now discuss the second part of Assumption D, viz., the condition, call it (M), that $\lim_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{n\kappa}(Z'Z/n)^+ \gamma_{n\kappa}) > 0$. Under Assumption A, condition (M) is equivalent to the requirement that the variance of $\Gamma_n(\hat{g})$ decreases as $n \rightarrow \infty$ at a rate no faster than $1/n$, which is not restrictive. Hence, in most cases, one only needs Assumption D to hold with the multiplicative factor $\kappa^{-1}(k)^{1/2}$ in (2.13).

As noted in Comment 7 following Theorem 1, condition (M) is automatically satisfied if $\Gamma_n(g) = g(x_j)$ for some x_j that is observed for n large and $\{z_s(\cdot)\}$ contains a constant function. Thus, it holds under Assumption F, provided $\{z_s(\cdot)\}$ contains a constant function. In addition, Theorem 2 of Section 3 below gives conditions under which condition (M) holds with probability one when the regressors are random.

For non-random regressors, sufficient conditions for condition (M) are: $\overline{\lim}_{n \rightarrow \infty} \lambda_{\max}(Z'_{n\kappa} Z_{n\kappa}) < \infty$, $Z_{n\kappa}$ is full rank κ_n for n large, and Assumption B(iii) holds.

Condition (M) is guaranteed to hold if $\kappa_n \leq \kappa_{bn}^*$ for n large, Assumption B(iii) holds, Assumption G holds, $D_1 = \sup_{k \geq 1} \overline{\lim}_{n \rightarrow \infty} \lambda_{\max}(Z'_{nk} Z_{nk}/n) < \infty$, and κ_{bn}^* is redefined as:

$$(2.18) \quad \kappa_{bn}^* = \max\{k \in I_+ : \max_{i \leq n} P Z_{nk}(i) \leq b(k), \max_{m \leq k} \gamma'_{nm} (Z'_{nm} Z_{nm})^+ \gamma_{nm} \leq b(k), \\ Z_{nk} \text{ is full rank } k, \text{ and } \lambda_{\max}(Z'_{nk} Z_{nk}/n) \leq 2D_1\}$$

and $\kappa_{bn}^* = 1$ if the set in (2.18) is empty. When $\Gamma_n(g) = g(x_j)$ for some observed x_j , the full rank condition in (2.18) is deleted. Under the assumptions, for each fixed k the conditions in (2.18) are met for n large, so $\kappa_{bn}^* \rightarrow \infty$ as $n \rightarrow \infty$.

We now discuss the main part of Assumption D. In almost all cases, the series functions $\{z_s(\cdot)\}$ are ordered such that higher values of s correspond to less smooth functions (for s large). In consequence, the smoother is g , the faster the coefficients die off in the series expansion of g in terms of $\{z_s(\cdot)\}$ and vice versa. Hence, Assumption D is a smoothness condition. Note that for any given κ there are an infinity of different functions $g \in \mathcal{G}$ such that Assumption D holds. Thus, Theorem 1(c) always has content—given κ , Assumption D holds if g is sufficiently smooth.

In some cases, it is possible to relate Assumption D to other smoothness conditions. For this purpose, we define the *Sobolev smoothness index* of g on the set $\mathcal{D} \subset \mathcal{X}$ to be

$$(2.19) \quad S_g(\mathcal{D}) = \max\{c \geq 0 : \|g\|_{c,\omega,\mathcal{D}} < \infty\}.$$

Suppose κ satisfies

$$(2.20) \quad \kappa_n \geq C n^{\omega_1} \text{ for some } \omega_1 \in (0,1) \text{ and some } C < \infty.$$

Then, $\kappa^{-1}(k) \leq C^* n^{1/\omega_1}$ for some $C^* < \infty$. Depending upon the regressor values and the choice of series functions, the satisfaction of (2.20) may or may not be consistent with κ satisfying Assumptions C and E. In Section 3 below we provide some results for random regressors where they are consistent with probability one. These results include (a) trigonometric series functions with $\mathcal{X} = [0, 2\pi]^d$ and regressors with (Lebesgue) density

bounded away from zero and (b) polynomial series functions with $\mathcal{X} = \prod_{r=1}^d [a_r, b_r]$ and regressors with (Lebesgue) density bounded below by some weight function.

Consider series functions from the FFF or trigonometric family and suppose \mathcal{X} is either $[0, 2\pi]^d$ or an open bounded set whose closure lies in $(0, 2\pi)^d$ and whose boundary is *minimally smooth* in the sense of Stein (1970, pp. 181, 189). (Alternatively, see Edmunds and Moscatelli (1977, p. 8) for the definition of minimally smooth.) When $\mathcal{X} = [0, 2\pi]^d$, assume g is periodic on $[0, 2\pi]$ in each coordinate. In these cases, Assumption D is satisfied if κ satisfies (2.20) and either

$$(2.21) \quad (i) S_g(\mathcal{D}) > q + \frac{d}{2\omega_1} \text{ or } (ii) S_g(\mathcal{D}) > q + \frac{d(\omega_1 + 1)}{2\omega_1},$$

where (i) applies if condition (M) holds and (ii) applies otherwise and $\mathcal{D} = \mathbb{R}^d$ if $\mathcal{X} = [0, 2\pi]^d$ and $\mathcal{D} = \mathcal{X}$ otherwise.⁵ Clearly, the periodicity condition used when $\mathcal{X} = [0, 2\pi]^d$ can be quite restrictive.

For trigonometric series, the result above follows from Edmunds and Moscatelli (1977, Cor. 1, p. 28). For FFF series, it follows from the trigonometric series result, since the addition of a finite number of non-trigonometric functions can only improve the approximation given by a trigonometric expansion without denigrating its rate.

As a second example, consider polynomial series functions defined on $\mathcal{X} = [a, b]$ for $-\infty < a < b < \infty$. In this case, Assumption D holds provided κ satisfies (2.20) and either

$$(2.22) \quad (i) S_g([a, b]) > q + \frac{1}{2\omega_1} \text{ or } (ii) S_g([a, b]) > q + \frac{\omega_1 + 1}{2\omega_1},$$

where (i) applies if condition (M) holds and (ii) applies otherwise. This result can be established using a polynomial approximation result given in Powell (1981, Thm. 3.2, p. 26). As stated, Powell's result only yields Assumption D for the case $q = 0$. His proof can be altered, however, to obtain the desired result for arbitrary integer $q \geq 0$. These results apply only when x is scalar. Analogous approximation results for multivariate

polynomials ($d > 1$) may exist, but are not known to the author.

Lastly, we note that for notational simplicity Assumptions A–G are concerned with a *sequence* of regressors $\{x_i : i \geq 1\}$. They can be generalized straightforwardly to the case of a *triangular array* of regressors $\{x_{ni} : i \leq n, n \geq 1\}$ and the results above and below and their proofs hold without change. Eastwood and Gallant's (1987) regression model, for example, is one in which the regressor variables form a triangular array.

3. RANDOM REGRESSORS

This section presents conditions under which (a) the key Assumptions C and E of Section 2 hold with probability one when the regressors are random and (b) explicit bounds on the growth rate of κ_n are obtained. Trigonometric and polynomial series functions and the estimands given in Examples 1–6 of (2.7) are treated in detail to exemplify the more general sufficient conditions given.

EXAMPLE I. This example considers trigonometric series functions defined on $[0, 2\pi]^d$ (see (2.2) for their definition). The random regressors $\{X_i : i \geq 1\}$ are assumed to be independent with (Lebesgue) densities $f_i(x)$ whose averages over $i = 1, \dots, n$ are bounded above and bounded away from zero for n large. That is, for some $N < \infty$,

$$(3.1) \quad 0 < \inf_{n \geq N} \inf_{x \in [0, 2\pi]^d} \frac{1}{n} \sum_{i=1}^n f_i(x) \leq \sup_{n \geq N} \sup_{x \in [0, 2\pi]^d} \frac{1}{n} \sum_{i=1}^n f_i(x) < \infty.$$

Example I is quite similar to Example 2 of Cox (1988, pp. 716, 717, 721, 722) except that Cox considers the case of a scalar regressor ($d = 1$) with uniform distribution.⁶ Cox's results are quite complementary to those of this paper, since he gives rates of convergence of derivative estimates in L_2 norms and establishes consistency in supremum norm, whereas we establish asymptotic normality.

EXAMPLE II. This example considers polynomial series functions on $\prod_{r=1}^d [a_r, b_r]$ ($= [a_1, b_1] \times \dots \times [a_d, b_d]$) for $-\infty < a_r < b_r < \infty$. That is, the k -vector of series functions $Z_k(x_i)$ is of the form

$$(3.2) \quad \begin{aligned} Z_k(x) &= S_k(Z_{k_1}^*(x_1) \otimes \dots \otimes Z_{k_d}^*(x_d)), \\ Z_{k_r}^*(x_r) &= (1, x_r, \dots, x_r^{k_r})', \quad \forall r = 1, \dots, d, x = (x_1, \dots, x_d)' \in \prod_{r=1}^d [a_r, b_r], \end{aligned}$$

for some positive integers k_1, \dots, k_d that satisfy $\prod_{r=1}^d k_r \geq k$, where S_k is a $k \times (\prod_{r=1}^d k_r)$ selection matrix for which $S_k S_k' = I_k$. We assume that the polynomial series functions are ordered such that every function in the $\prod_{r=1}^d (k_r - 1)$ -vector $Z_{k_1-1}^*(x_1) \otimes \dots \otimes Z_{k_d-1}^*(x_d)$ is included in the sequence before any of the additional functions in $Z_{k_1}^*(x_1) \otimes \dots \otimes Z_{k_d}^*(x_d)$ are included. In the univariate case, this means that $Z_k(x) = (1, x, \dots, x^k)'$. A consequence of this assumption is that $k \geq \prod_{r=1}^d (k_r/2)$, where k_1, \dots, k_d are the smallest integers for which $Z_k(\cdot)$ can be written as in (3.2).

In this example, the random regressors $\{X_i : i \geq 1\}$ are assumed to be independent with (Lebesgue) densities on $\prod_{r=1}^d [a_r, b_r]$ whose averages over $i = 1, \dots, n$ are bounded above and bounded below by some constant times a "beta" weight function for n large. In particular, we assume: For some $\epsilon > 0$, $N < \infty$, and $\alpha_r, \beta_r > -1 \quad \forall r = 1, \dots, d$,

$$\begin{aligned}
& \inf_{n \geq N} \frac{1}{n} \sum_{i=1}^n f_i(x) \geq \prod_{r=1}^d w_{\alpha_r, \beta_r}(x_r) \epsilon \quad \forall x = (x_1, \dots, x_d)' \in \prod_{r=1}^d [a_r, b_r] \text{ and} \\
(3.3) \quad & \sup_{n \geq N} \sup_{\substack{x \in \prod_{r=1}^d [a_r, b_r]}} \frac{1}{n} \sum_{i=1}^n f_i(x) < \infty, \text{ where} \\
& w_{\alpha_r, \beta_r}(x_r) = (b_r - x_r)^{\alpha_r} (x_r - a_r)^{\beta_r} \text{ for } r = 1, \dots, d.
\end{aligned}$$

When $a_r = 0$ and $b_r = 1 \quad \forall r$, $\prod_{r=1}^d w_{\alpha_r, \beta_r}(x_r)$ is proportional to the product of d beta densities. By taking $\alpha_r = \beta_r = 0 \quad \forall r$ one obtains a uniform weight function.

Example II is similar to Example 1 of Cox (1988, pp. 715, 716, 722–725) except that Cox considers the case of a scalar regressor on $[0,1]$ with beta distribution.⁷

We return now to the general case and introduce the following assumptions:

ASSUMPTION H: (i) $\{X_{ni} : i \leq n, n \geq 1\}$ is a triangular array of R^d -valued rv's that are independent across i for each fixed n .

(ii) For some $N < \infty$, $\inf_{n \geq N} \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n E Z_k(X_{ni}) Z_k(X_{ni})' \right] > 0 \quad \forall k \geq 1$.

(iii) $\sup_{x \in \mathcal{X}} |z_s(x)| < \infty \quad \forall s \geq 1$.

(iv) For some non-decreasing function $\alpha(\cdot) : I_+ \rightarrow [0, \infty)$, $\overline{\lim}_{n \rightarrow \infty} \max_{k \leq n} \lambda_{\max}(\gamma'_{nk} \gamma_{nk} / \alpha(k)) < \infty$.

ASSUMPTION I: Either (a) $\underline{\lim}_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{nk} \gamma_{nk}) > 0$ for some $k \geq 1$ and

$\overline{\lim}_{n \rightarrow \infty} \max_{k \leq n} \lambda_{\max} \left[\frac{1}{n} \sum_{i=1}^n E Z_k(X_{ni}) Z_k(X_{ni})' \right] < \infty$ or (b) $\underline{\lim}_{n \rightarrow \infty} \min_{k \leq n} \lambda_{\min}(\gamma'_{nk} \gamma_{nk}) / k > 0$

and $\sup_{x \in \mathcal{X}, s \geq 1} |z_s(x)| < \infty$.

None of the conditions in Assumptions H and I is restrictive except the independence requirement of Assumption H(i) and the nonsingularity requirement of Assumption H(ii). In fact, with m -dependent regressors the proof of Theorem 2 below goes through

with only minor alterations. In addition, the proof can be extended to cover strong mixing regressors without too much difficulty. Thus, independence is not crucial.

In the iid case, Assumption H(ii) requires $EZ_k(X_{ni})Z_k(X_{ni})'$ to be nonsingular for all $k \geq 1$. This is likely to be violated by FFF series for some k sufficiently large, since trigonometric functions are orthogonal and periodic on $[0, 2\pi]$ but the regressors are often scaled such that their support has closure contained in $(0, 2\pi)$ (see Gallant (1981, p. 217)) in order to avoid Gibb's phenomena at the endpoints. When Assumption H(ii) fails some series functions are redundant and need to be eliminated to make it hold.

Assumption H(iv) is quite weak. For example, if $\Gamma_n(\cdot)$ does not depend on n , then it is satisfied with $\alpha(k) = \max_{m \leq k} \lambda_{\max}(\gamma'_{nm} \gamma_{nm})$. In addition, it is always satisfied in Examples 1–6 of (2.7) provided Assumption H(iii) holds. In Examples 1–3 of (2.7), if Assumption H(iii) holds, then $\|\gamma_n(s)\|^2 \leq v \sup_{x \in \mathcal{X}} z_s^2(x) \forall s \geq 1$ and one can take

$$(3.4) \quad \alpha(k) = k \zeta^2(k) \quad \forall k \geq 1.$$

In Examples 4–6 of (2.7), $\|\gamma_n(s)\|^2 \leq v \max_{j \leq v} \sup_{x \in \mathcal{X}} |D^{\lambda_j} z_s(x)|^2$ and one can take

$$(3.5) \quad \alpha(k) = k \max_{s \leq k, j \leq v} \sup_{x \in \mathcal{X}} |D^{\lambda_j} z_s(x)|^2 \quad \forall k \geq 1.$$

Three functions $\zeta(\cdot)$, $\alpha(\cdot)$, and $\lambda(\cdot)$ are used in the main result of this section, i.e. Theorem 2 below. $\zeta(\cdot)$ is defined in (2.11) and $\alpha(\cdot)$ is defined in Assumption H(iv). Their values in Examples I and II are discussed below. When Assumption H(ii) holds $\lambda(\cdot)$ is defined to be a non-decreasing function from I_+ to $[1, \infty)$ such that for some $N < \infty$

$$(3.6) \quad \lambda(k) \inf_{n \geq N} \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n EZ_k(X_{ni})Z_k(X_{ni})' \right] \geq 1 \quad \forall k \geq 1.$$

The simplest choice of $\lambda(\cdot)$, of course, is $\lambda(k) = 1 / \inf_{n \geq N} \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n EZ_k(X_{ni})Z_k(X_{ni})' \right]$.

EXAMPLE I (cont.). Assumption H(i) holds in this example by assumption. Assumption H(ii) also holds, because for all $n \geq N$ and some $\delta, \tilde{\delta} \in (0,1]$,

$$(3.7) \quad \lambda_{\min} \left[\int_{[0,2\pi]^d} Z_k(x) Z_k(x)' \frac{1}{n} \sum_{i=1}^n f_i(x) dx \right] \geq \tilde{\delta} \lambda_{\min} \left[\int_{[0,2\pi]^d} Z_k(x) Z_k(x)' dx \right] \geq \delta$$

by the lower bound on $\frac{1}{n} \sum_{i=1}^n f_i(x)$ and the orthogonality of trigonometric functions on $[0, 2\pi]^d$. Thus, one can take $\lambda(k) = 1/\delta \quad \forall k \geq 1$. Since $\sup_{s \geq 1, x \in \mathcal{X}} |z_s(x)| \leq 1$, Assumption H(iii) holds and one can take $\zeta(k) = 1 \quad \forall k \geq 1$. Assumption H(iv) holds for the estimands of Examples 1–6 of (2.7) by the comments above regarding $\alpha(\cdot)$. In particular, one can take $\alpha(k) = k$ in Examples 1–3 by (3.4) and $\alpha(k) = k \frac{1+2\max_{j \leq v} |\lambda_j|}{d}$ in Examples 4–6 by (3.5) (when square, spherical, diamond, or equivalent partial sums are used in Examples 4–6, since in this case $\max_{j \leq v} \sup_{x \in \mathcal{X}} |D^{\lambda_j} z_s(x)| \leq C \max_{j \leq v} \prod_{m=1}^d (s^{1/d})^{\lambda_{jm}} = C s^{\max_{j \leq v} |\lambda_j|/d}$ for some $C < \infty$).

Assumption I(a) holds for Examples 1–6 of (2.7) in Example I provided no two of the vectors x_1, \dots, x_v are equal in Examples 2 or 4, no two of the vectors $\lambda_1, \dots, \lambda_v$ are equal in Example 5, and no two of the pairs $(\lambda_1, \eta_1), \dots, (\lambda_v, \eta_v)$ are equal in Example 6. This follows because with trigonometric series functions $\lim_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{nk} \gamma_{nk}) > 0$ for some $k \geq 1$ in Examples 1–6 under the provisos listed above and

$$(3.8) \quad \lambda_{\max} \left[\int_{[0,2\pi]^d} Z_k(x) Z_k(x)' \frac{1}{n} \sum_{i=1}^n f_i(x) dx \right] \leq \tilde{D} \lambda_{\max} \left[\int_{[0,2\pi]^d} Z_k(x) Z_k(x)' dx \right] \leq D < \infty$$

for all $n \geq N$ and some $D, \tilde{D} < \infty$, using the upper bound on $\frac{1}{n} \sum_{i=1}^n f_i(x)$.

EXAMPLE II (cont.). It is advantageous in this example to establish Assumptions C and E for orthonormalized polynomials rather than the simple polynomials given in (3.2). The shift to orthonormalized polynomials is innocuous because the replacement of $Z_{\kappa_n}(\cdot)$ by $G_{\kappa_n} Z_{\kappa_n}(\cdot)$ in the calculation of \hat{g} , where G_{κ_n} is a nonsingular $\kappa_n \times \kappa_n$ matrix, leaves \hat{g} , and hence $\Gamma_n(\hat{g})$, unchanged. Thus, for theoretical purposes only, we can suppose that $Z_k(\cdot)$ is as defined in (3.2) but with

$$\begin{aligned} Z_{k_r}^*(x_r) &= (z_1^*(x_r), \dots, z_{k_r}^*(x_r))', \\ (3.9) \quad z_s^*(x_r) &= c_{\alpha_r, \beta_r}^{(s-1)} P_{s-1}^{(\alpha_r, \beta_r)}(2(x_r - a_r)/(b_r - a_r) - 1), \text{ and} \\ c_{\alpha_r, \beta_r}^{(s-1)} &= \left[\frac{1}{b_r - a_r} \right]^{(1+\alpha_r+\beta_r)/2} \left[\frac{(2s-2+\alpha_r+\beta_r+1)\Gamma(s)\Gamma(s+\alpha_r+\beta_r)}{\Gamma(s+\alpha_r)\Gamma(s+\beta_r)} \right]^{1/2} \end{aligned}$$

for $r = 1, \dots, d$ and $s = 1, 2, \dots$, where $\{P_{s-1}^{(\alpha_r, \beta_r)}(x) : s \geq 1\}$ are the Jacobi polynomials on $[-1, 1]$ with parameters (α_r, β_r) , see Szegő (1939, Ch. 4) or Abramowitz and Stegun (1964, p. 775, eqn. 22.3.1), and $\Gamma(\cdot)$ is the gamma function.

If Assumptions C and E are verified using the orthonormalized polynomials of (3.9), then so must the other assumptions in A–E that involve the series functions, viz., Assumptions B(iii) and D. This is not a problem, because Assumption B(iii) is implied by Assumption I (with both defined using the orthonormalized polynomials) and Assumption D holds with the orthonormalized polynomials of (3.9) if and only if it holds with the simple polynomials of (3.2).

As shown in the Appendix (following the proof of Theorem 1), the orthonormal polynomials of (3.9) satisfy Assumptions H(ii) and H(iii) with $\lambda(\cdot)$ and $\zeta(\cdot)$ given by

$$(3.10) \quad \lambda(k) = 1 \quad \forall k \geq 1 \text{ and}$$

$$(3.11) \quad \zeta(k) = Ck^\xi \text{ for } \xi = \max\{\alpha_1 + 1/2, \beta_1 + 1/2, \dots, \alpha_d + 1/2, \beta_d + 1/2, 0\}$$

for some $C < \infty$. By (3.4), Assumption H(iv) holds in Examples 1–3 of (2.7) with

$$(3.12) \quad \alpha(k) = k\zeta^2(k) = Ck^{1+2\xi} \quad \forall k \geq 1.$$

By (3.5) and calculations in the Appendix, Assumption H(iv) holds in Examples 4–6 of (2.7) with

$$(3.13) \quad \alpha(k) = k^{1+2\nu} \quad \forall k \geq 1, \text{ where} \\ \nu = \max \left\{ \max \left[\alpha_r + 2\lambda_{jr} + \frac{1}{2}, \beta_r + 2\lambda_{jr} + \frac{1}{2}, \lambda_{jr} \right] : r = 1, \dots, d; j = 1, \dots, v \right\}.$$

Assumption I(a) holds for Examples 1–6 of (2.7) in Example II (using the orthonormalized polynomials of (3.9)) under the same provisos as given in Example I above and for the same reasons.

We now state the main result of this section:

THEOREM 2: (a) *Under Assumption H, if κ is nondecreasing, $\kappa_n \rightarrow \infty$ as $n \rightarrow \infty$, and*

$$(3.14) \quad \max\{\kappa_n^4 \lambda^2(\kappa_n) \zeta^4(\kappa_n), \lambda(\kappa_n) \alpha(\kappa_n)\} / n \leq Dn^{-\tau} \text{ for all } n \text{ large,}$$

for some $0 < \tau < 1$ and some $D < \infty$, then Assumptions C and E hold with probability one when the regressors $\{X_{ni}\}$ are random.

(b) *Under Assumptions H and I, if κ is as above, then $\liminf_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{n\kappa}(Z'Z/n)^+ \gamma_{n\kappa}) \geq C$ for some constant $C > 0$ with probability one when the regressors $\{X_{ni}\}$ are random. In addition, Assumption B(iii) holds.*

COMMENTS: 1. For any given functions $\lambda(\cdot)$, $\zeta(\cdot)$, and $\alpha(\cdot)$, truncation sequences κ that satisfy the conditions of Theorem 2 always exist. For example, the truncation sequence $\tilde{\kappa}$ satisfies these conditions, where

$$(3.15) \quad \tilde{\kappa}_n = \max \left\{ k \in I_+ : \max\{k^4 \lambda^2(k) \zeta^4(k), \lambda(k) \alpha(k)\} / n \leq Dn^{-\tau} \right\}$$

and $\tilde{\kappa}_n = 1$ if the set above is empty. More importantly, if $\lambda(\cdot)$, $\zeta(\cdot)$, and $\alpha(\cdot)$ are bounded by powers of k , then truncation sequences κ exist that satisfy the conditions of

Theorem 2 and grow as fast as a power of n .

2. If $\lambda(\cdot)$ and $\zeta(\cdot)$ are bounded above and $\alpha(k) \leq \tilde{C}k^4$ for some $\tilde{C} < \infty$, then (3.14) requires $\kappa_n \leq Dn^\omega$ for $\omega < 1/4$ and some $D < \infty$. This is the fastest rate allowable in Theorem 2. When these conditions do not hold, the maximum rate is slower. These maximum rates may be restrictive, especially when d is large, and it would be desirable to increase them. On the other hand, they are much faster than some rates that appear in the literature. For example, Newey (1988a) requires his polynomial series estimators to satisfy $\kappa_n \ln \kappa_n = o(\ln n)$ (whereas the polynomial series estimators of Example II may have κ_n grow as a power of n , see below).

3. The proof of Theorem 2 shows that if $\lambda(\cdot)$ is bounded, as it is in Examples I and II, then $\lim_{n \rightarrow \infty} \lambda_{\min}(Z'_{n\kappa} Z_{n\kappa}/n) > 0$ with probability one under the assumptions of Theorem 2. (See Lemma A-1(b) in the Appendix.) This result is useful for several purposes, including showing that series estimators of the parameters in semiparametric regression models are \sqrt{n} -consistent, see Section 5.2 below.

EXAMPLE I (cont.). Suppose the errors $\{U_i\}$ satisfy Assumption A conditional on $\{X_i\} = \{x_i\}$ with probability one. (For example, this holds if the errors are iid, square integrable, and independent of the regressors.) From above, $\zeta(k) = \delta\lambda(k) = 1$ and $\alpha(k) = k \quad \forall k \geq 1$. Thus, condition (3.14) of Theorem 2 requires $\kappa_n \leq Dn^\omega$ for some $\omega < 1/4$ and $D < \infty$. Under this condition and the others stated above for Example I, Assumptions C and E hold conditional on $\{X_i\}$ with probability one. In addition, for Examples 1-6 of (2.7), Assumption B holds with $q = 0, 0, 0, |\lambda|, \max_{j \leq v} |\lambda_j|$, and $\max_{j \leq v} |\lambda_j|$ respectively. Thus, the result of Theorem 1(a) holds conditionally on $\{X_i\}$ with probability one and unconditionally.

If, in addition to the conditions above, g is periodic in each of its elements,

$$(3.16) \quad D_1 n^{\omega_1} \leq \kappa_n \leq D_2 n^{\omega_2} \text{ for } 0 < \omega_1 \leq \omega_2 < 1/4 \text{ for } n \text{ large, and}$$

$$S_g(R^d) > q + \frac{d}{2\omega_1}$$

for some $D_1, D_2 < \infty$, then Assumptions A–E hold conditional on $\{X_i\}$ with probability one for Examples 1–6 of (2.7) and the results of Theorem 1(b) and (c) hold conditional on $\{X_i\}$ with probability one and unconditionally (using (2.21) to verify Assumption D).

To contrast the asymptotic normality results here with the mean squared error results of Cox (1988), we note that in Cox's Example 2 with uniform regressor distribution κ_n has an upper bound of $D_2 n^{\omega_2}$ for some $\omega_2 < 1/3$ and $D_2 < \infty$. (See Cox (1988, p. 722) and set $p = 1/2$ in the equation following his equation (3.6).) Thus, his results allow a slightly faster growth rate of κ_n than does Theorem 2.

EXAMPLE II (cont.). Suppose the errors $\{U_i\}$ satisfy Assumption A conditional on $\{X_i\} = \{x_i\}$ with probability one. From above, $\lambda(k) = 1$ and $\zeta(k) = Ck^\xi$ for ξ as in (3.11) for all $k \geq 1$. For the estimands of Examples 1–3 of (2.7), $\alpha(k) = C^2 k^{1+2\xi}$ for all $k \geq 1$. Thus, for Examples 1–3 of (2.7), condition (3.14) of Theorem 2 requires $\kappa_n \leq Dn^\omega$ for some $\omega < 1/(4+4\xi)$. For example, for the uniform weight function (i.e., $\alpha_1 = \beta_1 = \dots = \beta_d = 0$), we need $\omega < 1/6$. Under the above condition on κ_n and the other conditions stated above for Example II, Assumptions C and E hold conditionally on $\{X_i\}$ with probability one for Examples 1–3 of (2.7). In addition, Assumption B holds with $q = 0$. Hence, the result of Theorem 1(a) holds conditionally on $\{X_i\}$ with probability one and unconditionally.

If, in addition to the conditions above, we have a scalar regressor ($d = 1$),⁸

$$(3.17) \quad D_1 n^{\omega_1} \leq \kappa_n \leq D_2 n^{\omega_2} \text{ for } 0 < \omega_1 \leq \omega_2 < 1/(4+4\xi) \text{ for } n \text{ large, and}$$

$$S_g([a,b]) > \frac{1}{2\omega_1}$$

for some $D_1, D_2 < \infty$, then Assumptions A–E hold conditional on $\{X_i\}$ with probability one for Examples 1–3 of (2.7) and the results of Theorem 1(b) and (c) hold conditional on $\{X_i\}$ with probability one and unconditionally (using (2.22) to verify Assumption D).

For Examples 4–6 of (2.7), $\alpha(k) = k^{1+2\nu}$ for ν as in (3.13). Thus, condition (3.14) of Theorem 2 holds in this case if $\kappa_n \leq Dn^\omega$ for $\omega < 1/\max\{4+4\xi, 1+2\nu\}$. If $\max_{r \leq d, j \leq v} \lambda_{jr} \leq 1$ (i.e., the estimand only involves the 0–th and 1–st order derivatives of g), then this upper bound on ω equals $1/(4+4\xi)$ as in Examples 1–3. For Examples 4–6 of (2.7), the results of the preceding paragraph hold provided $1/(4+4\xi)$ is replaced by $1/\max\{4+4\xi, 1+2\nu\}$ in (3.17).

In contrast to the results above, Cox (1988) requires $\kappa_n \leq D_2 n^\omega$ for $\omega < 1/(3+4\xi)$ for his polynomial series estimator. (See Cox (1988, p. 715, Ex. 1) and set $p = 1/2$ and $h = \xi + 1/2$ in the equation preceding his equation (1.2).) For example, with the uniform weight function, Cox requires $\omega < 1/5$ whereas Theorem 2 requires $\omega < 1/6$. As mentioned above, Newey (1988a) requires $\kappa_n \ln \kappa_n = o(\ln n)$ for his polynomial series estimators, which corresponds to a much smaller upper bound on κ_n than is needed here.

4. RANDOM TRUNCATION SEQUENCES

In this section we extend the asymptotic normality results of Section 2 to series estimators that are based on certain data-dependent truncation sequences. The need for this extension is due to the fact that a good choice of κ_n depends on the unknown function g . In consequence, one needs to use information in the data about g when choosing κ_n .

Let $\hat{\kappa}$ denote the random truncation sequence used to form the estimator \hat{g} . That is, \hat{g} is as in (2.5) but with $\hat{\kappa}$ in place of κ . Let Z abbreviate $Z_{n\hat{\kappa}}$. Define

$$(4.1) \quad V_n^* = \gamma_{n\hat{\kappa}}'(Z'Z)^+ Z' \Omega Z (Z'Z)^+ \gamma_{n\hat{\kappa}} \text{ and } A_n^* = (V_n^*)^{-1/2}.$$

Note that V_n^* and A_n^* are random, whereas V_n and A_n are non-random.

We require $\hat{\kappa}$ to satisfy an assumption that is quite similar to assumptions of

Eastwood (1987) and Eastwood and Gallant (1987), but is more general. (It encompasses two alternative assumptions regarding $\hat{\kappa}$ that these authors consider.) Nevertheless, the assumption is still restrictive and its relaxation would be desirable.

Let \mathcal{K} denote a *countable* collection of truncation sequences κ^j for $j \geq 1$ such that Assumptions C and E hold for each $\kappa^j \in \mathcal{K}$. Using Theorem 1(d) or (e) or Theorem 2, it is possible to specify a class \mathcal{K} that satisfies these conditions without great difficulty.

The random truncation sequence $\hat{\kappa}$ is assumed to satisfy:

ASSUMPTION J: $\hat{\kappa}_n - \tilde{\kappa}_n \xrightarrow{P} 0$ as $n \rightarrow \infty$ for some random truncation sequence $\tilde{\kappa}$ that takes values in \mathcal{K} .

Examples of random truncation sequences $\hat{\kappa}$ that satisfy Assumption J are Eastwood and Gallant's (1987) and Eastwood's (1987, 1988) upward F-testing rules, where the F-tests test upward from a random truncation sequence $\tilde{\kappa}$ that is countably-valued. See the references above for details. Other examples of random truncation sequences that satisfy Assumption J include those that only depend on discretized (i.e., rational-valued) versions of the first N observations for some $N < \infty$ (even as the sample size n goes to infinity).

THEOREM 3: Under Assumptions A, B, D, and J, the series estimator $\{\Gamma_n(\hat{g})\}$ based on the random truncation sequence $\hat{\kappa}$ satisfies

$$A_n^*(\Gamma_n(\hat{g}) - \Gamma_n(g)) \rightarrow N(0, I_V) \text{ as } n \rightarrow \infty \text{ and}$$

$$\Gamma_n(\hat{g}) - \Gamma_n(g) \xrightarrow{P} 0 \text{ as } n \rightarrow \infty.$$

COMMENT: Random truncation sequences that are common in the literature include those generated by C_p , cross-validation, and generalized cross-validation. See Li (1987) and Andrews (1989) for analyses of the asymptotic optimality properties of these procedures with respect to the minimization of average squared error. Assumption J and Theorem 3 do not apply to these truncation rules. In fact, the asymptotic normality result

of Theorem 3 cannot be expected to hold for these procedures because they do not drive the ratio of the bias to the standard deviation of $\Gamma_n(\hat{g})$ to zero.

5. AIR, SEMIPARAMETRIC REGRESSION, AND INDEX MODELS

This section discusses the application of the results of Sections 2–4 to models other than the fully nonparametric regression model considered above. AIR, semiparametric regression, and semiparametric index regression models are discussed.

5.1. Additive Interactive Regression Models

This section considers series estimators of *additive interactive regression* (AIR) models. These models are also known in the literature as interaction spline models.⁹ They have been analyzed using spline estimators by Barry (1983, 1986), Wahba (1986), Gu, Bates, Chen, and Wahba (1988), and Chen (1988). A special case of the AIR model is the *additive regression* model that has been considered by Orcutt *et al.* (1961, p. 243), Stone (1985), Hastie and Tibshirani (1986, 1987), and Buja, Hastie, and Tibshirani (1989).

By definition, an *AIR model of order two* is a nonparametric regression model in which the unknown regression function g is of the form

$$(5.1) \quad g(x_i) = \sum_{j=1}^d g_j(x_{ij}) + \sum_{j=1}^d \sum_{m=j+1}^d g_{jm}(x_{ij}, x_{im}),$$

where g_j and g_{jm} are unknown functions and $x_i = (x_{i1}, \dots, x_{id})'$.¹⁰ AIR models of order ρ for integers $\rho \geq 1$ are defined analogously but with interactive functions that include up to ρ regressor variables. Of course, an AIR model of order d is equivalent to a fully nonparametric regression model.

AIR models lie between linear regression models and fully nonparametric regression models in terms of the restrictions they place on the regression functions. Their attractiveness is due to the fact that they are considerably more general than linear regression models, but they circumvent the "curse of dimensionality" that afflicts the estimation of

fully nonparametric regression models. In particular, it is shown in Andrews and Whang (1989) that the rate of convergence to zero of the mean average squared error of series estimators of AIR models of order ρ (defined below) with regressors of arbitrary dimension is the same as that of series estimators of fully nonparametric regression models with regressors of dimension ρ . Similar results for spline estimators (but under quite restrictive assumptions on the regressors) are obtained by Chen (1988).¹¹

The AIR models described above are *saturated* AIR models. Unsaturated AIR models, in which interactions are introduced between some regressors but not others, also may be useful. This is especially true, of course, if the modelling context suggests which interactions should be included and which should be excluded. If d is "large" relative to n , an unsaturated AIR model of order two may be as general a nonparametric regression model as one can estimate with reasonable accuracy.

The estimation of additive regression and AIR models via series estimators is straightforward. One merely takes a set of functions $\{z_s(\cdot) : s \geq 1\}$, which is suitable for estimating a fully nonparametric regression function, and eliminates those functions $z_s(\cdot)$ whose coefficients in an expansion of g in terms of $\{z_s(\cdot)\}$ must necessarily equal zero for any function g that is of the form specified by the model. For example, for an FFF series estimator of a second order AIR model, the functions $z_s(\cdot)$ are of the form given in (2.2) but the vectors $b \in \mathbb{R}^d$ that are used are restricted to vectors that have at most two non-zero elements. If the original sequence of series functions is dense with respect to the supremum Sobolev norm in a set \mathcal{G} , then the new sequence is dense in the subset of functions in \mathcal{G} that satisfy the model.

A convenient property of the results of Sections 2–4 for fully nonparametric regression models is that they apply in the present context as well. To apply them, one makes the same definitions and assumptions as in Sections 2–4 except the family of approximating functions $\{z_s(\cdot)\}$ is restricted as described in the paragraph above. As long as the sequence $\{z_s(\cdot)\}$ approximates g sufficiently well to satisfy Assumption D, there is no

difference between applying Theorems 1–3 to AIR models than to fully nonparametric regression models. In consequence, under the assumptions of Theorems 1 or 3, series estimators of additive regression and AIR models are consistent and asymptotically normal. Consistent estimators of their covariance matrices are provided in Section 6 below.

5.2. Semiparametric Regression Models

Next, we consider a semiparametric regression model. This model is of the form:

$$(5.2) \quad Y_i = x_{ai}'\beta + g_b(x_{bi}) + U_i \text{ for } i = 1, \dots, n,$$

where Y_i , x_{ai} , and x_{bi} are observed, $\{U_i\}$ are unobserved, independent, mean zero errors, and β and g_b are unknown.

This model has been considered by Robinson (1988), Chamberlain (1986, 1987), Wahba (1986), and Rice (1986) among others. See Engle *et al.* (1986) and Stock (1985) for empirical applications of this model. Note that in addition to being used as models in their own right, semiparametric regression models can be used in constructing tests for omitted variables in fully nonparametric or linear regression models.

Robinson (1988) and Chamberlain (1986) are concerned with \sqrt{n} -consistent estimation of β in the model (5.2). Here, we are interested in estimating both β and various functions of g_b , such as those given in the examples of (2.7).

As in Chamberlain (1986), we consider estimating the regression function $g(x) = x_a'\beta + g_b(x_b)$ by a series estimator that takes account of the special structure of g . In particular, the family of functions used to approximate g is taken to include the elements of x_a plus a family of functions of x_b that approximate g_b .

The results of Sections 2–4 can be applied in the present context. To do so, one makes the same definitions and assumptions as in Sections 2–4 except that the family of series functions is defined as above. The parameter vector β can be estimated by taking $\Gamma_n(g) = \frac{\partial}{\partial x_a}g(x)$ for any $x \in \mathcal{X}$, since $\frac{\partial}{\partial x_a}g(x) = \frac{\partial}{\partial x_a}(x_a'\beta + g_b(x_b)) = \beta$. This yields $\hat{\beta}$

to be the vector of LS estimators of the coefficients on x_a from the regression of Y_i on x_{ai} and certain functions of x_{bi} . In addition to estimation of β , by appropriate choice of $\Gamma_n(\cdot)$, the same functions of g_b can be estimated as in Sections 2–4. Assumption D is satisfied in the present case if it is satisfied with g replaced by g_b .

A consequence of the remarks above is that Theorems 1 and 3 establish the consistency and asymptotic normality of the estimator $\Gamma_n(\hat{g})$ of $\Gamma_n(g)$ in the semiparametric regression context and Section 6 below establishes the consistency of estimators of the covariance matrix of $\Gamma_n(\hat{g})$ in this model. In the homoskedastic variance case, the covariance matrix estimator for $\hat{\beta}$ is just the standard LS covariance matrix estimator for a parameter sub-vector in a linear regression model.

An important question is whether an estimator of β in (5.2) can achieve the \sqrt{n} rate of consistency. Robinson (1988) establishes this result for a kernel-based LS estimator and Chamberlain (1986) does likewise for a series estimator. In the present context, it is not difficult to see that the series estimator $\hat{\beta}$ is \sqrt{n} consistent and asymptotically normal if $\lim_{n \rightarrow \infty} \lambda_{\min}(Z'_{nk} Z_{nk}/n) > 0$. For example, if Assumption G holds with Assumption G(ii) replaced by $\lim_{n \rightarrow \infty} \lambda_{\min}(Z'_{nk} Z_{nk}/n) \geq \epsilon \quad \forall k \geq 1$ for some $\epsilon > 0$ and the condition $\lambda_{\min}(Z'_{nk} Z_{nk}/n) \geq \epsilon/2$ is added to the definition of κ_b^* , then the sufficient condition $\lim_{n \rightarrow \infty} \lambda_{\min}(Z'_{nk} Z_{nk}/n) > 0$ holds. For iid regressors, the condition above holds with probability one if $\lim_{n \rightarrow \infty} \lambda_{\min}(EZ_k(X_i)Z_k(X_i)') \geq \epsilon \quad \forall k \geq 1$ for some $\epsilon > 0$. Alternatively, for inid regressors, $\lim_{n \rightarrow \infty} \lambda_{\min}(Z'_{nk} Z_{nk}/n) > 0$ with probability one if Assumption H and (3.14) hold and $\lambda(\cdot)$ is bounded above (see Comment 3 following Theorem 2).

We note that the results given here cover the case of heteroskedastic errors, as occurs in sample selection examples among others. Robinson's (1988) results do not cover this case, but he conjectures that they can be extended in this direction.

5.3. Semiparametric Index Regression Models

A *single index regression (SIR) model* for the random variables $\{Y_i, x_i : 1 \leq i \leq n\}$ is any model in which

$$(5.3) \quad E(Y_i | x_i) = F(x_i' \beta) \text{ for } i = 1, \dots, n,$$

where $\beta \in R^d$ is unknown and $F(\cdot) : R \rightarrow R$ may be known, known up to a finite dimensional parameter, or unknown. If $F(\cdot)$ is unknown, then the model is called a *semiparametric SIR model*. Examples of semiparametric SIR models are numerous. They include a wide variety of limited dependent variable models, such as binary choice models, in which latent errors have unspecified distributions. See Ruud (1986) and Stoker (1986).

A semiparametric SIR model is a special case of the nonparametric regression model of (2.1): If (5.3) holds, then one can always write

$$(5.4) \quad Y_i = F(x_i' \beta) + U_i, \quad i = 1, \dots, n,$$

where the errors $\{U_i\}$ have conditional mean zero given $\{x_i\}$ and may be heteroskedastic. In the notation of Section 2, $g(\cdot) = F((\cdot)' \beta)$. For present purposes, we assume that (Y_i, x_i) (and hence, U_i) is independent across observations.

Following Stoker (1986), we note that the derivative with respect to x of the conditional mean of Y_i is proportional to β in SIR models.¹² Thus, if $\Gamma_n(\cdot)$ is chosen as in Example 5 of (2.7), then

$$(5.5) \quad \Gamma_n(g) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial x} g(x_i) = \left[\frac{1}{n} \sum_{i=1}^n F_1(x_i' \beta) \right] \beta,$$

$$(5.6) \quad \Gamma_n(\hat{g}) = \gamma'_{n\kappa} \hat{\theta}_{n\kappa} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial}{\partial x'} Z_{\kappa_n}(x_i) \right]' \hat{\theta}_{n\kappa},$$

and the vector $\Gamma_n(g)$ is proportional to β , where $F_1(\cdot)$ denotes the scalar derivative of $F(\cdot)$. Under the assumptions of Theorem 1 or 3, $\Gamma_n(\hat{g})$ is consistent for β up to scale and asymptotically normal.¹³ Consistent estimators of its covariance matrix are provided in Section 6. Computationally, $\Gamma_n(\hat{g})$ is quite simple, because it is just a linear

combination of linear least squares estimators.

Alternative estimators of the semiparametric SIR model that are known to be asymptotically normal include those of Powell, Stock, and Stoker (1989), Ichimura (1986), and Härdle and Stoker (1987). In fact, these estimators are also known to be \sqrt{n} -consistent, which has not been established for the series estimators considered here.

The estimators of Powell *et al.* (1989) and Härdle and Stoker (1986) also can be applied to the estimation of various sorts of weighted averages derivatives. For example, Härdle and Stoker (1986) consider estimating

$$(5.7) \quad \Gamma_n(g) = \int_X \frac{\partial}{\partial x} g(x) dH(x),$$

where the regressors $\{X_i\}$ are random with unknown distribution function H . The results of Sections 2–4 do not apply to this estimand, but they do apply to the estimation of in-sample average derivatives (see Example 5 of (2.7)) and to the estimation of weighted average derivatives with known weight functions (see Example 6 of (2.7)).

A generalization of the semiparametric SIR model is the *semiparametric multiple index regression (MIR) model*. For rv's $\{Y_i, x_i : 1 \leq i \leq n\}$, such models are characterized by

$$(5.8) \quad E(Y_i | x_i) = F(x'_{1i}\beta_1, x'_{2i}\beta_2, \dots, x'_{pi}\beta_p) \text{ for } i = 1, \dots, n,$$

for some partition $(x'_{1i}, \dots, x'_{pi})'$ of x_i , some unknown parameters $(\beta'_1, \dots, \beta'_p)'$, and some unknown function $F : R^p \rightarrow R$. Multinomial discrete choice and sample selection models based on latent linear regression models with unknown error distributions are examples of semiparametric MIR models. See Stoker (1986) for further discussion. The extension of the results above to the estimation of the parameters $(\beta_1, \dots, \beta_p)$ of the semiparametric MIR model is straightforward.

6. COVARIANCE MATRIX ESTIMATION

In this section we consider estimation of the covariance matrix V_n or V_n^* of the estimator $\Gamma_n(\hat{g})$. The results apply to estimators based on random and non-random truncation sequences. They apply in nonparametric regression, additive regression, AIR, semiparametric regression, and semiparametric index regression models. The two cases of homoskedastic and heteroskedastic error variances are treated.

First, we consider the case of homoskedastic errors. The following assumption applies when the truncation sequence $\hat{\kappa}$ is random. It is imposed in conjunction with Assumption J of Section 4.

ASSUMPTION K: (i) $\sigma_i^2 = \sigma^2 > 0$ for $i = 1, 2, \dots$.
(ii) $\sup_{i \geq 1} E|U_i|^{2+\delta} < \infty$ for some $\delta > 0$.
(iii) For each $\kappa^j \in \mathcal{K}$, $\kappa_n^j \leq \bar{\kappa}_n$ for all n , for some non-random truncation sequence $\bar{\kappa}$ that satisfies $\bar{\kappa}_n \max_{i \leq n} PZ_{n\bar{\kappa}}(i) \rightarrow 0$ as $n \rightarrow \infty$.

A truncation sequence $\bar{\kappa}$ satisfies $\bar{\kappa}_n \max_{i \leq n} PZ_{n\bar{\kappa}}(i) \rightarrow 0$ as $n \rightarrow \infty$ under Assumption F or G with κ replaced by $\bar{\kappa}$, provided κ_b and κ_b^* are defined with $PZ_{n\kappa}(i)$ replaced by $kPZ_{n\bar{\kappa}}(i)$. Alternatively, for random regressors, a truncation sequence $\bar{\kappa}$ satisfies $\bar{\kappa}_n PZ_{n\bar{\kappa}}(i) \rightarrow 0$ as $n \rightarrow \infty$ with probability one if $\bar{\kappa}$ satisfies (3.14) and Assumption H holds. (This is proved analogously to the proof of Theorem 2(a) using Lemma A-1(a) of the Appendix.)

The case of a non-random truncation sequence κ is covered by Assumption K by taking \mathcal{K} to have only one element, κ , and by taking $k = \bar{\kappa} = \kappa$. Then, Assumption K(iii) reduces to $\kappa_n \max_{i \leq n} PZ_{n\kappa}(i) \rightarrow 0$ as $n \rightarrow \infty$.

Under Assumption K, the estimator of V_n^* is given by

$$(6.1) \quad \hat{V}_n = \gamma'_{n\hat{\kappa}} (Z'_{n\hat{\kappa}} Z_{n\hat{\kappa}})^+ \gamma_{n\hat{\kappa}} \hat{\sigma}^2$$

where $\hat{\sigma}^2 = \frac{1}{n - \hat{\kappa}_n} \sum_{i=1}^n (Y_i - \hat{g}(x_i))^2$. Correspondingly, the estimator \hat{A}_n of A_n^* is given by $\hat{V}_n^{-1/2}$. Consistency of these estimators is established in the following theorem:

THEOREM 4: *Suppose the series estimators $\{\Gamma_n(\hat{g})\}$ are based on a random truncation sequence $\hat{\kappa}$. Then, under Assumptions A, B, D, J, and K, we have $\hat{V}_n(V_n^*)^{-1} \xrightarrow{P} I_V$, $\hat{A}_n(A_n^*)^{-1} \xrightarrow{P} I_V$, and $\hat{A}_n(\Gamma_n(\hat{g}) - \Gamma_n(g)) \xrightarrow{d} N(0, I_V)$ as $n \rightarrow \infty$.*

Next, we consider the case of heteroskedastic errors. As above, we only treat estimators based on random truncation sequences $\hat{\kappa}$. Results for estimators based on non-random truncation sequences κ are obtained as a special case. Five different estimators of V_n^* and A_n^* are considered. For $w = 1, \dots, 5$, let

$$\begin{aligned} \hat{V}_{wn} &= \gamma'_{n\hat{\kappa}} (\hat{Z}'\hat{Z})^+ \hat{Z}' \hat{\Omega}_{wn} \hat{Z} (\hat{Z}'\hat{Z})^+ \gamma_{n\hat{\kappa}} \text{ and } \hat{A}_{wn} = \hat{V}_{wn}^{-1/2}, \text{ where} \\ (6.2) \quad \hat{\Omega}_{1n} &= \text{diag}(\hat{U}_i^2), \quad \hat{\Omega}_{2n} = \text{diag}(\hat{U}_i^2 n / (n - \hat{\kappa}_n)), \quad \hat{\Omega}_{3n} = \text{diag}(\hat{U}_i^2 / (1 - z_{ii})), \\ \hat{\Omega}_{4n} &= \text{diag}\left[(n-1)\hat{U}_i^2 / (n(1 - z_{ii})^2)\right] - \frac{n-1}{n^2} \tilde{U} \tilde{U}', \text{ and } \hat{\Omega}_{5n} = \text{diag}\left[\hat{U}_i^2 / (1 - z_{ii})^2\right], \\ \text{and where } \hat{Z} &\text{ denotes } Z_{n\hat{\kappa}}, \quad \hat{U}_i = Y_i - \hat{g}(x_i), \quad z_{ii} \text{ denotes } PZ_{n\hat{\kappa}}(i), \quad \tilde{U}_i = \hat{U}_i / (1 - z_{ii}), \\ \text{and } \tilde{U} &= (\tilde{U}_1, \dots, \tilde{U}_n)'. \end{aligned}$$

The estimators $\hat{V}_{1n}, \dots, \hat{V}_{4n}$ are analogues of estimators that have been suggested in the literature for estimating the covariance matrix of the LS estimator in the standard linear regression model. \hat{V}_{1n} is an analogue of an estimator proposed by Eicker (1963a, 1967) and popularized in the econometrics literature by White (1980). \hat{V}_{2n} is a variant of \hat{V}_{1n} due to Hinckley (1977) that incorporates a simple degrees of freedom adjustment. \hat{V}_{3n} is a variant of \hat{V}_{1n} proposed by Horn, Horn, and Duncan (1975) that incorporates an alternative bias reduction adjustment. \hat{V}_{4n} is the jack-knife estimator of V_n^* (see equation (12) of MacKinnon and White (1985)).

The estimator \hat{V}_{5n} is the analogue of the delete-one cross-validation estimator of the covariance matrix of the LS estimator in the standard linear regression model. It has

not been considered previously in the literature (to my knowledge). It has the attribute that it is never downward biased (for any regressors and any Ω) in the latter model, and hence, t-statistics based on it over-reject less often than t-statistics based on $\hat{V}_{1n}, \dots, \hat{V}_{4n}$.

For each $\kappa^j \in \mathcal{K}$, let V_n^j denote $\gamma'_{n\kappa^j} (Z'Z)^+ Z' \Omega Z (Z'Z)^+ \gamma_{n\kappa^j}$, where Z denotes $Z_{n\kappa^j}$. V_n^j is the covariance matrix of $\Gamma_n(\hat{g})$ when $\Gamma_n(\hat{g})$ is based on κ_n^j terms.

To establish the consistency of \hat{V}_{wn} and \hat{A}_{wn} , we assume:

ASSUMPTION L: (i) $\sup_{i \geq 1} EU_i^4 < \infty$.

(ii) Assumption K(iii) holds.

(iii) $b_n^j V_n^j \rightarrow V_j$ as $n \rightarrow \infty$ for some constants $\{b_n^j : n \geq 1\}$ and some positive definite $v \times v$ matrix V_j , for each $j \geq 1$ that corresponds to a truncation sequence κ^j in \mathcal{K} .

THEOREM 5: Suppose the estimators $\{\Gamma_n(\hat{g})\}$ are based on a random truncation rule $\hat{\kappa}$. Then, under Assumptions A, B, D, J, and L, we have $\hat{V}_n(V_n^*)^{-1} \xrightarrow{P} I_v$, $\hat{A}_n(A_n^*)^{-1} \xrightarrow{P} I_v$, and $\hat{A}_n(\Gamma_n(\hat{g}) - \Gamma_n(g)) \xrightarrow{d} N(0, I_v)$ as $n \rightarrow \infty$.

COMMENT: MacKinnon and White (1985) and Davidson and MacKinnon (1985) report results from a number of Monte Carlo experiments comparing the performance of t-statistics constructed using $\hat{V}_{1n}, \dots, \hat{V}_{4n}$ in the standard linear regression model. They find that \hat{V}_{4n} performs best in terms of the closeness of the true and nominal sizes of the tests based on the t-statistics. Similar simulation results by the author show that \hat{V}_{5n} outperforms \hat{V}_{4n} using the same criterion by a small but not insignificant margin.¹⁴

APPENDIX

This Appendix contains proofs of the results stated in the text.

PROOF OF THEOREM 1: First we establish part (a). By Assumptions A(ii), B(iii), and C(iii), V_n is nonsingular and A_n is well-defined for n sufficiently large. To see this, note that either Z is full rank κ_n , in which case

$$(A.1) \quad \lambda_{\min}(V_n) \geq \inf_{i \geq 1} \sigma_i^2 \lambda_{\min}(\gamma'_{n\kappa} (Z'Z)^{-1} \gamma_{n\kappa}) \geq \inf_{i \geq 1} \sigma_i^2 \lambda_{\min}(\gamma'_{n\kappa} \gamma_{n\kappa}) / \lambda_{\max}(Z'Z) > 0$$

for n large, or $\Gamma_n(g) = g(x_j)$, in which case

$$(A.2) \quad \lambda_{\min}(V_n) \geq \inf_{i \geq 1} \sigma_i^2 \lambda_{\min}(e'_j \bar{Z} (\bar{Z}' \bar{Z})^{-1} \bar{Z}' e_j) \geq \inf_{i \geq 1} \sigma_i^2 z_s^2(x_j) / \lambda_{\max}(\bar{Z}' \bar{Z}) > 0$$

for n large, where e_j is the j -th elementary n -vector, s is as in Assumption C(iii), and \bar{Z} is an $n \times \bar{\kappa}_n$ submatrix of Z of full rank $\bar{\kappa}_n$.

Consider any $b \in R^V$ with $\|b\| = 1$. Let $\eta = Z(Z'Z)^+ \gamma_{n\kappa} A_n b \in R^n$. To establish part (a), it suffices to show that $\eta' U \xrightarrow{d} N(0,1)$ as $n \rightarrow \infty$. By the central limit theorem of Eicker (1963b, Thm. 1), this holds if (i) $\{U_i : i \geq 1\}$ are mean zero iid rv's, (ii) $\inf_{i \geq 1} \sigma_i^2 > 0$, (iii) $\sup_{i \geq 1} E U_i^2 1(|U_i| > c) \rightarrow 0$ as $c \rightarrow \infty$, and (iv) $\max_{i \leq n} \eta_i^2 / \eta' \eta \rightarrow 0$ as $n \rightarrow \infty$. Conditions (i)–(iii) hold by Assumptions A(i)–A(iii) respectively. Condition (iv) holds if $\max_{i \leq n} |\eta_i| \rightarrow 0$ as $n \rightarrow \infty$ and $\lim_{n \rightarrow \infty} \eta' \eta > 0$. We have

$$(A.3) \quad \lim_{n \rightarrow \infty} \eta' \eta \geq \lim_{n \rightarrow \infty} \sum_{i=1}^n \sigma_i^2 \eta_i^2 / \sup_{j \geq 1} \sigma_j^2 = 1 / \sup_{j \geq 1} \sigma_j^2 > 0$$

using $\eta' \Omega \eta = 1$ and Assumption A(ii). In addition, we have

$$\begin{aligned}
\max_{i \leq n} |\eta_i| &\leq \max_{i \leq n} \left[Z_{\kappa_n}(x_i)' (Z'Z)^+ Z_{\kappa_n}(x_i) \right]^{1/2} \left[b' A_n \gamma'_{n\kappa} (Z'Z)^+ \gamma_{n\kappa} A_n b \right]^{1/2} \\
(A.4) \quad &\leq \max_{i \leq n} (PZ_{n\kappa}(i))^{1/2} \left[\sum_{m=1}^n \eta_m^2 \sigma_m^2 / \inf_{j \geq 1} \sigma_j^2 \right]^{1/2} \\
&\leq \max_{i \leq n} (PZ_{n\kappa}(i))^{1/2} \left[\eta' \Omega \eta / \inf_{j \geq 1} \sigma_j^2 \right]^{1/2} \rightarrow 0 \text{ as } n \rightarrow \infty
\end{aligned}$$

by Assumptions A(ii) and C(ii). This establishes part (a).

To prove part (b), we show that for any $b \in R^V$ with $\|b\| = 1$

$$(A.5) \quad Q_n = b' A_n (E\Gamma_n(\hat{g}) - \Gamma_n(g)) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Let $g_k = Z_k(\cdot)' \theta_k$, where θ_k is as in Assumption D. Let g_k^I be the remainder function from approximating g by g_k : $g_k^I = g - g_k$. For any $\tilde{g} \in \mathcal{G}$, let $\tilde{g}(X)$ denote $(\tilde{g}(x_1), \dots, \tilde{g}(x_n))'$. Note that $g_k(X) = Z_{n\kappa} \theta_k$. By Assumption C(iii), $\gamma'_{n\kappa} (Z'Z)^+ (Z'Z) = \gamma'_{n\kappa}$ for n large. Thus, for n large we have

$$\begin{aligned}
E\Gamma_n(\hat{g}) &= \gamma'_{n\kappa} (Z'Z)^+ Z' g_{\kappa_n}(X) + \gamma'_{n\kappa} (Z'Z)^+ Z' g_{\kappa_n}^I(X) \\
&= \gamma'_{n\kappa} (Z'Z)^+ Z' Z \theta_{\kappa_n} + \gamma'_{n\kappa} (Z'Z)^+ Z' g_{\kappa_n}^I(X) \\
&= \gamma'_{n\kappa} \theta_{\kappa_n} + \gamma'_{n\kappa} (Z'Z)^+ Z' g_{\kappa_n}^I(X), \\
(A.6) \quad &
\end{aligned}$$

$$\Gamma_n(g) = \gamma'_{n\kappa} \theta_{\kappa_n} + \Gamma_n(g_{\kappa_n}^I), \quad Q_n = Q_{1n} - Q_{2n},$$

$$Q_{1n} = b' A_n \gamma'_{n\kappa} (Z'Z)^+ Z' g_{\kappa_n}^I(X), \text{ and } Q_{2n} = b' A_n \Gamma_n(g_{\kappa_n}^I).$$

To show $Q_{1n} \rightarrow 0$ as $n \rightarrow \infty$, we write

$$\begin{aligned}
|Q_{1n}| &\leq \left[b' A_n \gamma'_{n\kappa} (Z'Z)^+ \gamma_{n\kappa} A_n b \right]^{1/2} \left[g_{\kappa_n}^r(X)' Z (Z'Z)^+ Z' g_{\kappa_n}^r(X) \right]^{1/2} \\
(A.7) \quad &\leq \left[b' A_n \gamma'_{n\kappa} (Z'Z)^+ Z' \Omega Z (Z'Z)^+ \gamma_{n\kappa} A_n b / \inf_{i \geq 1} \sigma_i^2 \right]^{1/2} n^{1/2} \max_{i \leq n} |g_{\kappa_n}^r(x_i)| \\
&\leq \left[1 / \inf_{i \geq 1} \sigma_i^2 \right]^{1/2} \kappa_n^{-1} (\kappa_n)^{1/2} \|g - g_{\kappa_n}\|_{q, \omega, \mathcal{X}} \rightarrow 0 \text{ as } n \rightarrow \infty,
\end{aligned}$$

where the convergence to zero uses Assumptions A(ii), C(i), and D.

To show $Q_{2n} \rightarrow 0$ as $n \rightarrow \infty$, we write

$$(A.8) \quad Q_{2n} \leq \lambda_{\max}(A_n) \|\Gamma_n(g_{\kappa_n}^r)\| \leq [1/\lambda_{\min}(V_n)]^{1/2} C_1 \|g_{\kappa_n}^r\|_{q, \omega, \mathcal{X}}$$

for n large, using Assumption B(ii). In addition, we have

$$(A.9) \quad \lambda_{\max}(Z'Z) = \sup_{c: \|c\|=1} \sum_{i=1}^n (c' Z_{\kappa_n}(x_i))^2 \leq n \sum_{s=1}^{\kappa_n} \sup_{x \in \mathcal{X}} z_s^2(x) \leq n \kappa_n \zeta^2(\kappa_n).$$

Thus, when Z is full rank κ_n , by (A.1), (A.8), (A.9) and Assumptions B(iii), C(i), and D, we have: For some constant $C < \infty$,

$$(A.10) \quad Q_{2n} \leq C \left[\kappa_n^{-1} (\kappa_n) \kappa_n \right]^{1/2} \zeta(\kappa_n) \|g_{\kappa_n}^r\|_{q, \omega, \mathcal{X}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

If $\lim_{n \rightarrow \infty} \lambda_{\min}(\gamma'_{n\kappa} (Z'Z/n)^+ \gamma_{n\kappa}) > 0$, then the right hand side of the first line of (A.1) is bounded below by $\inf_{i \geq 1} \sigma_i^2 C_2/n$ for all n large for some $C_2 > 0$ and $\left[\kappa_n^{-1} (\kappa_n) \kappa_n \right]^{1/2} \zeta(\kappa_n)$ can be replaced by $\kappa_n^{-1} (\kappa_n)^{1/2}$ in (A.10). In consequence, $Q_{2n} \rightarrow 0$ in this case with this same replacement made in Assumption D. When Z is not of full rank κ_n , (A.2) plus $\lambda_{\max}(\bar{Z}'\bar{Z}) \leq \lambda_{\max}(Z'Z)$ must be used in place of (A.1) to get (A.10).

To show part (c) of the Theorem, first suppose Assumption E holds. Then, $\forall j = 1, \dots, v$, $\text{Var}(e_j' \Gamma_n(\hat{g})) \leq e_j' \gamma'_{n\kappa} (Z'Z)^+ \gamma_{n\kappa} e_j \sup_{i \geq 1} \sigma_i^2 \rightarrow 0$ as $n \rightarrow \infty$, using Assumptions A(ii) and E, where e_j is the j -th elementary v -vector. This result and (A.5) imply

that the mean squared error of $\Gamma_n(\hat{g})$ goes to zero as $n \rightarrow \infty$.

Next, suppose $\Gamma_n(\hat{g}) - \Gamma_n(g) \xrightarrow{P} 0$ as $n \rightarrow \infty$ and Assumption E does not hold. Then, $V_n \not\rightarrow 0$ as $n \rightarrow \infty$ and for some sequence of constants $\{c_n \in \mathbb{R}^V : n \geq 1\}$ with $\|c_n\| = 1 \forall n$, $c_n' A_n^2 c_n \not\rightarrow \infty$ as $n \rightarrow \infty$. Thus, we have

$$(A.11) \quad |c_n' A_n (\Gamma_n(\hat{g}) - \Gamma_n(g))| \leq \left[c_n' A_n^2 c_n \right]^{1/2} \|\Gamma_n(\hat{g}) - \Gamma_n(g)\| \xrightarrow{P} 0$$

as $n \rightarrow \infty$, which contradicts the result of Theorem 1 part (b). In consequence, Assumption E is a necessary condition for weak consistency of $\Gamma_n(\hat{g})$ under Assumptions A-D. Since $\text{Var}(b' \Gamma_n(\hat{g}) b) \geq \inf_{i \geq 1} \sigma_i^2 b' \gamma_{n\kappa} (Z'Z)^+ \gamma_{n\kappa} b$ for $b \in \mathbb{R}^V$ with $\|b\| = 1$, Assumption E is clearly a necessary condition for L^2 consistency of $\Gamma_n(\hat{g})$.

Now, we prove part (d). Under Assumption F(i), Assumption C implies Assumption E, because

$$(A.12) \quad \gamma_{n\kappa}' (Z'Z)^+ \gamma_{n\kappa} = e_j' Z (Z'Z)^+ Z' e_j = PZ_{n\kappa}(j)$$

for n large. Also, Assumptions F(i) and (ii) imply Assumptions C(iii) and (i) respectively. Thus, it remains to show Assumption C(ii). For fixed k , if $\int Z_k(x) Z_k(x)' dF(x)$ is nonsingular, then $Z_{nk}' Z_{nk} / n$ is nonsingular for n large and

$$(A.13) \quad \begin{aligned} \max_{i \leq n} PZ_{nk}(i) &\leq \frac{1}{n} \max_{i \leq n} \|Z_{nk}' e_i\|^2 / \lambda_{\min}(Z_{nk}' Z_{nk} / n) \\ &\leq \frac{1}{n} \sum_{s=1}^k \sup_{x \in \mathcal{X}} z_s^2(x) / \lambda_{\min}(Z_{nk}' Z_{nk} / n) \rightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

using Assumptions F(iii) and (v), where e_i denotes the i -th elementary n -vector.

If $\int Z_k(x) Z_k(x)' dF(x)$ is singular, we remove the minimal number of series functions from the k -vector $Z_k(\cdot)$ that yields a nonsingular integral. (If a nonsingular integral cannot be obtained by removing functions from $Z_k(\cdot)$, then it must be the case that $\int (c' Z_k(x))^2 dF(x) = 0 \forall c \in \mathbb{R}^k$, $Z_k(x_i) = 0 \forall i \geq 1$ by Assumption F(iv), and $PZ_{nk}(i) = 0 \forall i$, $\forall n$, as desired.) The same functions can be removed from Z_{nk} with-

out affecting the value of $\max_{i \leq n} PZ_{nk}(i)$, because Assumption F(iv) implies that the functions removed correspond to redundant columns of Z_{nk} and $\max_{i \leq n} PZ_{nk}(i)$ is a function of Z_{nk} only through the space spanned by the columns of Z_{nk} . Now, the same argument as in (A.13) applied to the transformed Z_{nk} matrices yields $\max_{i \leq n} PZ_{nk}(i) \rightarrow 0$ as $n \rightarrow \infty$. The latter result implies that the set in (2.14) is non-empty for n large. Assumption C(ii) follows, because (2.14) implies that

$$(A.14) \quad \max_{i \leq n} PZ_{n\kappa}(i) \leq b(\kappa_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Lastly, we establish part (e) of the Theorem. Assumptions G(i) and (ii) imply Assumptions C(i) and (iii) respectively. Assumptions G(ii) and (iii) give: For n large,

$$(A.15) \quad \begin{aligned} \max_{i \leq n} PZ_{nk}(i) &\leq \max_{i \leq n} \|Z'_{nk} e_i\|^2 / \lambda_{\min}(Z'_{nk} Z_{nk}) \\ &\leq \sum_{s=1}^k \sup_{x \in \mathcal{X}} z_s^2(x) / \lambda_{\min}(Z'_{nk} Z_{nk}) \rightarrow 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

Assumptions G(ii) and (iv) give: For n large,

$$(A.16) \quad \lambda_{\max}(\gamma'_{nk} (Z'_{nk} Z_{nk})^{-1} \gamma_{nk}) \leq \lambda_{\max}(\gamma'_{nk} \gamma_{nk}) / \lambda_{\min}(Z'_{nk} Z_{nk}) \rightarrow 0$$

as $n \rightarrow \infty$. (A.15), (A.16), and Assumption G(ii) imply that the set in (2.15) is non-empty for n large. In consequence, (A.14) holds and

$$(A.17) \quad \lambda_{\max}(\gamma'_{n\kappa} (Z'Z)^+ \gamma_{n\kappa}) \leq b(\kappa_n) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus, Assumptions C(ii) and E hold. \square

We now compute the functions $\lambda(\cdot)$, $\zeta(\cdot)$, and $\alpha(\cdot)$ for Example II of Section 3. The constants $c_{\alpha_r, \beta_r}^{(s-1)}$ in (3.9) have been chosen to normalize the given (shifted and rescaled) sequences of Jacobi polynomials (see Szegő (1939, p. 67, eqn. 4.3.1) or Abramowitz and Stegun (1964, p. 774, eqn. 22.2.1)):

$$(A.18) \quad \int_{[a_r, b_r]} z_s^*(x_r) z_t^*(x_r) w_{\alpha_r, \beta_r}(x_r) dx_r = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases} \quad \forall r = 1, \dots, d.$$

Since $\lambda_{\min}(\text{SDS}') \geq \lambda_{\min}(\text{SS}') \lambda_{\min}(\text{D})$ and $\lambda_{\min}(\text{D}_1 \bullet \text{D}_2) = \lambda_{\min}(\text{D}_1) \lambda_{\min}(\text{D}_2)$ for conformable matrices S and D and positive semidefinite matrices D_1 and D_2 , equation (A.18) yields

$$(A.19) \quad \begin{aligned} & \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n \text{E} \text{S}_{\mathbf{k}} \text{Z}_{\mathbf{k}}(\text{X}_i) \text{Z}_{\mathbf{k}}(\text{X}_i)' \text{S}_{\mathbf{k}}' \right] \\ & \geq \lambda_{\min}(\text{S}_{\mathbf{k}} \text{S}_{\mathbf{k}}') \prod_{r=1}^d \lambda_{\min} \left[\int_{[a_r, b_r]} \text{Z}_{\mathbf{k}_r}^*(x_r) \text{Z}_{\mathbf{k}_r}^*(x_r)' \frac{1}{n} \sum_{i=1}^n f_i(x_r) dx_r \right] \\ & \geq \prod_{r=1}^d \lambda_{\min} \left[\int_{[a_r, b_r]} \text{Z}_{\mathbf{k}_r}^*(x_r) \text{Z}_{\mathbf{k}_r}^*(x_r)' w_{\alpha_r, \beta_r}(x_r) dx_r \epsilon \right] = \epsilon > 0 \end{aligned}$$

for all $n \geq N$ and all $k \geq 1$. In consequence, Assumption H(ii) holds and we can take $\lambda(\cdot)$ as in (3.10).

We now establish Assumption H(iii) and determine $\zeta(\cdot)$ for the orthonormalized polynomials of (3.9). Standard results for the gamma function (see Abramowitz and Stegun (1964, p. 257, eqn. 6.1.46)) yield

$$(A.20) \quad c_{\alpha_r, \beta_r}(s-1) \sim \sqrt{s} \quad (\text{i.e., } c_{\alpha_r, \beta_r}(s-1)/\sqrt{s} \rightarrow 1 \text{ as } s \rightarrow \infty).$$

In addition, standard inequalities for Jacobi polynomials (see Szegő (1939, p. 163, eqn. 7.32.2) or Abramowitz and Stegun (1964, p. 786, eqn. 22.14.1)) yield

$$(A.21) \quad \sup_{x \in [-1, 1]} |P_{s-1}^{(\alpha, \beta)}(x)| \sim s^{\max\{\alpha, \beta, -1/2\}}$$

provided $\alpha, \beta > -1$. By the assumption concerning the ordering of the polynomials,

$k \geq \prod_{r=1}^d k_r/2$. These results combine to give

$$\begin{aligned}
(A.22) \quad & \max_{s \leq k} \sup_{\substack{x \in X \\ r=1}}^d |z_s(x)| \leq \sup_{\substack{x \in X \\ r=1}}^d \max_{s \leq k} \prod_{r=1}^d |z_{s_r}^*(x_r)| \\
& \leq C^* \prod_{r=1}^d k_r^{h+1/2} \leq C^*(2k)^{h+1/2}
\end{aligned}$$

for some $C^* < \infty$, where $h = \max\{\alpha_1, \beta_1, \dots, \alpha_d, \beta_d, -1/2\}$ (and s_r denotes the order of the polynomial in x_r whose product over r yields $z_s(x)$). Hence, Assumption H(iii) holds and we can take $\zeta(\cdot)$ as in (3.11).

To determine $\alpha(\cdot)$ for Examples 4-6 of (2.7) using (3.5), we need a bound on $\max_{s \leq k, j \leq v} \sup_{\substack{x \in X \\ r=1}}^d \left| D^{\lambda_j} z_s(x) \right|^2$ in terms of k . By (4.21.7) of Szegő (1939, p. 62),

$$(A.23) \quad \frac{d}{dx} P_{s-1}^{(\alpha, \beta)}(x) = \frac{1}{2}(s + \alpha + \beta) P_{s-2}^{(\alpha+1, \beta+1)}(x).$$

Combined with (A.21) this gives

$$(A.24) \quad \sup_{x \in [-1, 1]} \left| \frac{d^t}{dx^t} P_{s-1}^{(\alpha, \beta)}(x) \right| \leq \tilde{C}_s \max\{\alpha+2t, \beta+2t, t-1/2\} \text{ for } t = 0, 1, \dots,$$

for some $\tilde{C} < \infty$. Thus, we obtain

$$\begin{aligned}
(A.25) \quad & \sup_{\substack{x \in X \\ r=1}}^d |D^{\lambda_j} z_s(x)| \leq \sup_{\substack{x \in X \\ r=1}}^d \prod_{r=1}^d \left| \frac{d^{\lambda_{jr}}}{dx_r^{\lambda_{jr}}} z_{s_r}^*(x_r) \right| \\
& = \sup_{\substack{x \in X \\ r=1}}^d \prod_{r=1}^d c_{\alpha_r, \beta_r}(s_r - 1) \left| \frac{d^{\lambda_{jr}}}{dx_r^{\lambda_{jr}}} P_{s_r-1}^{(\alpha_r, \beta_r)} \left[2 \frac{x_r - a_r}{b_r - a_r} - 1 \right] \right| \\
& \leq C^* \prod_{r=1}^d s_r^{1/2} \max\{\alpha_r + 2\lambda_{jr}, \beta_r + 2\lambda_{jr}, \lambda_{jr} - 1/2\}
\end{aligned}$$

for some $C^* < \infty$. By the assumption regarding the ordering of the polynomials,

$s > \prod_{r=1}^d (s_r/2)$, and hence,

$$(A.26) \quad \sup_{\substack{x \in \bigcap_{r=1}^d [a_r, b_r]}} |D^{\lambda_j}_{z_s}(x)| \leq C^{**} s^{\nu}$$

for ν as in (3.13) and for some $C^{**} < \infty$. In consequence, for Example II with the estimands of Examples 4-6 of (2.7) we can take $\alpha(\cdot)$ as in (3.13).

Next we prove Theorem 2. The proof uses the following lemma:

LEMMA A-1: Let κ be as in Theorem 2. Then, for random regressors $\{X_{ni}\}$,

(a) Assumption H implies

$$\lambda_{\min} \left[\frac{1}{\max\{\kappa_n^4 \lambda(\kappa_n) \zeta^4(\kappa_n), \alpha(\kappa_n)\}} \sum_{i=1}^n Z_{\kappa_n}(X_{ni}) Z_{\kappa_n}(X_{ni})' \right] \rightarrow \infty \text{ a.s.},$$

(b) Assumption H and $\lambda(\cdot)$ bounded imply

$$\lim_{n \rightarrow \infty} \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n Z_{\kappa_n}(X_{ni}) Z_{\kappa_n}(X_{ni})' \right] > 0 \text{ a.s., and}$$

(c) Assumptions H and I(a) imply

$$\overline{\lim}_{n \rightarrow \infty} \lambda_{\max} \left[\frac{1}{n} \sum_{i=1}^n Z_{\kappa_n}(X_{ni}) Z_{\kappa_n}(X_{ni})' \right] < \infty \text{ a.s.}$$

PROOF OF THEOREM 2: First we prove part (a). By assumption, Assumption C(i) holds. By Lemma A-1, Assumption C(iii) holds with probability one. To establish Assumption C(ii) we write: For n large,

$$(A.27) \quad \begin{aligned} \max_{i \leq n} P Z_{n\kappa}(i) &= \max_{i \leq n} e_i' Z_{n\kappa} (Z_{n\kappa}' Z_{n\kappa})^{-1} Z_{n\kappa}' e_i \\ &\leq \max_{i \leq n} \|Z_{n\kappa}' e_i\|^2 / \lambda_{\min}(Z_{n\kappa}' Z_{n\kappa}) \leq \sum_{s=1}^{\kappa_n} \sup_{x \in \mathcal{X}} z_s^2(x) / \lambda_{\min}(Z_{n\kappa}' Z_{n\kappa}) \\ &\leq \kappa_n \zeta^2(\kappa_n) / \lambda_{\min}(Z_{n\kappa}' Z_{n\kappa}) \rightarrow 0 \text{ as } n \rightarrow \infty \text{ a.s.} \end{aligned}$$

using Lemma A-1(a).

To show that Assumption E holds, we write:

$$(A.28) \quad \begin{aligned} \overline{\lim}_{n \rightarrow \infty} \lambda_{\max} \left[\gamma'_{n\kappa} (Z'_{n\kappa} Z_{n\kappa})^{-1} \gamma_{n\kappa} \right] &\leq \overline{\lim}_{n \rightarrow \infty} \lambda_{\max} (\gamma'_{n\kappa} \gamma_{n\kappa} / \alpha(\kappa_n)) \alpha(\kappa_n) / \lambda_{\min} (Z'_{n\kappa} Z_{n\kappa}) \\ &\leq \overline{\lim}_{n \rightarrow \infty} \max_{k \leq n} \lambda_{\max} (\gamma'_{nk} \gamma_{nk} / \alpha(k)) \alpha(\kappa_n) / \lambda_{\min} (Z'_{n\kappa} Z_{n\kappa}) = 0 \text{ a.s.} \end{aligned}$$

using Assumption H(iv) and Lemma A-1(a).

Next, we establish part (b) of the Theorem. Suppose Assumption I(a) holds. Then,

$$(A.29) \quad \underline{\lim}_{n \rightarrow \infty} \lambda_{\min} (\gamma'_{n\kappa} (Z' Z / n)^{-1} \gamma_{n\kappa}) \geq \underline{\lim}_{n \rightarrow \infty} \lambda_{\min} (\gamma'_{n\kappa} \gamma_{n\kappa}) / \lambda_{\max} (Z' Z / n) > 0 \text{ a.s.}$$

using Lemma A-1(c).

Alternatively, suppose Assumption I(b) holds. Then,

$$(A.30) \quad \lambda_{\max} (Z'_{n\kappa} Z_{n\kappa} / (n\kappa_n)) = \sup_{c: \|c\|=1} \frac{1}{n\kappa_n} \sum_{i=1}^n (c' Z_{\kappa_n} (X_{ni}))^2 \leq \frac{1}{\kappa_n} \sum_{s=1}^{\kappa_n} \sup_{x \in \mathcal{X}} z_s^2(x) < \infty$$

using the Cauchy-Schwartz inequality. Thus, we get

$$(A.31) \quad \underline{\lim}_{n \rightarrow \infty} \lambda_{\min} (\gamma'_{n\kappa} (Z' Z / n)^+ \gamma_{n\kappa}) \geq \underline{\lim}_{n \rightarrow \infty} \min_{k \leq n} \lambda_{\min} (\gamma'_{nk} \gamma_{nk} / k) / \lambda_{\max} (Z' Z / (n\kappa_n)) > 0$$

using (A.30) and Assumption I(b).

Assumption B(iii) follows immediately from Assumption I(a) or I(b). \square

PROOF OF LEMMA A-1: First, we establish parts (a) and (b). For any two psd $k \times k$ matrices A and B , let c be the unit k -vector defined by $c' B c = \lambda_{\min}(B)$ if $\lambda_{\min}(A) > \lambda_{\min}(B)$ and $c' A c = \lambda_{\min}(A)$ otherwise. Then,

$$(A.32) \quad \begin{aligned} |\lambda_{\min}(A) - \lambda_{\min}(B)| &\leq |c' A c - c' B c| \\ &= \left| \sum_{s=1}^k \sum_{t=1}^k (a_{st} - b_{st}) c_s c_t \right| \leq \sum_{s=1}^k \sum_{t=1}^k |a_{st} - b_{st}|. \end{aligned}$$

Let X_i abbreviate X_{ni} . Using (A.32) and Markov's inequality, we get

$$\begin{aligned}
& P \left[\left| \lambda_{\min} \left[\frac{\lambda(\kappa_n)}{n} \sum_{i=1}^n Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right] - \lambda_{\min} \left[\frac{\lambda(\kappa_n)}{n} \sum_{i=1}^n E Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right] \right| < \epsilon \right] \\
& \leq P \left[\sum_{s=1}^{\kappa_n} \sum_{t=1}^{\kappa_n} \left| \frac{\lambda(\kappa_n)}{n} \sum_{i=1}^n (z_s(X_i) z_t(X_i) - E z_s(X_i) z_t(X_i)) \right| > \epsilon \right] \\
& \leq \frac{1}{\epsilon} \sum_{s=1}^{\kappa_n} \sum_{t=1}^{\kappa_n} E \left| \frac{\lambda(\kappa_n)}{n} \sum_{i=1}^n (z_s(X_i) z_t(X_i) - E z_s(X_i) z_t(X_i)) \right| \\
& \leq \frac{1}{\epsilon} \sum_{s=1}^{\kappa_n} \sum_{t=1}^{\kappa_n} \left[\frac{\lambda^2(\kappa_n)}{n^2} \sum_{i=1}^n E (z_s(X_i) z_t(X_i) - E z_s(X_i) z_t(X_i))^2 \right]^{1/2} \\
& \leq \frac{1}{\epsilon} \left[\kappa_n^4 \lambda^2(\kappa_n) \zeta^4(\kappa_n) / n \right]^{1/2}.
\end{aligned}
\tag{A.33}$$

The "in probability" versions of Lemma A-1(a) and (b) follow from (A.33), because the last line of (A.33) goes to zero as $n \rightarrow \infty$ by (3.14), $\lambda_{\min} \left[\frac{\lambda(\kappa_n)}{n} \sum_{i=1}^n E Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right] > 0$ by definition of $\lambda(\cdot)$ (using Assumption H(ii)), and $\left[1 / \max \{ \kappa_n^4 \lambda(\kappa_n) \zeta^4(\kappa_n), \alpha(\kappa_n) \} \right] / (\lambda(\kappa_n)/n) \rightarrow \infty$ by (3.14).

To obtain the "almost sure" results of Lemma A-1(a) and (b), however, a more complicated argument is needed. For any two truncation sequences κ_1 and κ_2 , define

$$\text{(A.34)} \quad \xi(\kappa_1, \kappa_2, n) = \frac{\lambda(\kappa_{1n})}{n} \lambda_{\min} \left[\sum_{i=1}^n Z_{\kappa_{2n}}(X_i) Z_{\kappa_{2n}}(X_i)' \right].$$

We introduce a new truncation sequence $\check{\kappa}$ and an infinite subsequence $\{n_m : m \geq 1\}$ of $\{n : n \geq 1\}$ such that

$$\begin{aligned}
& \text{(i)} \quad \check{\kappa}_n \geq \kappa_n \quad \forall n \geq 1, \\
& \text{(ii)} \quad \check{\kappa} \text{ satisfies condition (3.14) of Theorem 2,} \\
& \text{(A.35)} \quad \text{(iii)} \quad n_m / n_{m+1} \rightarrow 1 \text{ as } m \rightarrow \infty, \\
& \text{(iv)} \quad \lim_{m \rightarrow \infty} \xi(\check{\kappa}, \check{\kappa}, n_m) > 0 \text{ a.s., and} \\
& \text{(v)} \quad \min_{n \in [n_m, n_{m+1}]} \xi(\check{\kappa}, \kappa, n) \geq \frac{n_m}{n_{m+1}} \xi(\check{\kappa}, \check{\kappa}, n_m) \quad \forall m \geq 1.
\end{aligned}$$

Properties (iii)–(v) yield:

$$(A.36) \quad \lim_{n \rightarrow \infty} \xi(\check{\kappa}, \kappa, n) = \lim_{m \rightarrow \infty} \min_{n \in [n_m, n_{m+1}]} \xi(\check{\kappa}, \kappa, n) \geq \lim_{m \rightarrow \infty} \left[\frac{n_m}{n_{m+1}} \right] \xi(\check{\kappa}, \check{\kappa}, n_m) > 0 \text{ a.s.},$$

which establishes part (b) of the Lemma. In addition, properties (i) and (ii) yield:

$$(A.37) \quad \left[\frac{1}{\max\{\kappa_n^4 \lambda(\kappa_n) \zeta^4(\kappa_n), \alpha(\kappa_n)\}} \right] / \left[\frac{\lambda(\check{\kappa}_n)}{n} \right] \geq \frac{n}{\max\{\check{\kappa}_n^4 \lambda^2(\check{\kappa}_n) \zeta^4(\check{\kappa}_n), \lambda(\check{\kappa}_n) \alpha(\check{\kappa}_n)\}} \rightarrow \infty$$

as $n \rightarrow \infty$. Combining (A.36) and (A.37) establishes part (a) of the Lemma.

For parts (a) and (b), it remains to define $\check{\kappa}$ and $\{n_m\}$ and to prove (i)–(v) of (A.35). Let p be an integer greater than $2/\tau$. Define

$$(A.38) \quad \check{\kappa}_n = \kappa_{[(n^{1/p} + 1)^p]} \quad \forall n \geq 1 \text{ and } n_m = m^p \quad \forall m \geq 1,$$

where $[\cdot]$ denotes the integer part of \cdot . Property (i) holds immediately from the definition of $\check{\kappa}$ and the fact that κ is non-decreasing. In fact, $\check{\kappa}$ satisfies the stronger property

$$(A.39) \quad \check{\kappa}_{m^p} = \kappa_{(m+1)^p} \geq \kappa_n \quad \forall n \leq (m+1)^p, \quad \forall m \geq 1,$$

which is used below to establish property (v).

To establish property (ii), note that for all n greater than 2^p and large enough that (3.14) holds, we have

$$(A.40) \quad \begin{aligned} \max\{\kappa_n^4 \lambda^2(\kappa_n) \zeta^4(\kappa_n), \lambda(\kappa_n) \alpha(\kappa_n)\} &\leq D [n^{1/p}]^{p(1-\tau)} \leq D [2n^{1/p} - 2]^{p(1-\tau)} \\ &= D^* [n^{1/p} - 1]^{p(1-\tau)}, \end{aligned}$$

where $D^* = D 2^{p(1-\tau)}$. Replacing n by $[(n^{1/p} + 1)^p]$ in (A.40) yields

$$(A.41) \quad \max\{\check{\kappa}_n^4 \lambda^2(\check{\kappa}_n) \zeta^4(\check{\kappa}_n), \lambda(\check{\kappa}_n) \alpha(\check{\kappa}_n)\} \leq D^* n^{1-\tau},$$

which establishes property (ii).

Property (iii) follows immediately from the definition of $\{n_m\}$.

Property (iv) is established using (A.33), the Borel–Cantelli Lemma, and Assumption H(ii). Note that (A.33) holds with κ replaced by $\check{\kappa}$ throughout. In addition,

$$(A.42) \quad \sum_{m=1}^{\infty} \left[\check{\kappa}_{n_m}^4 \lambda^2(\check{\kappa}_{n_m}) \zeta^4(\check{\kappa}_{n_m}) / n_m \right]^{1/2} \leq D^* \sum_{m=1}^{\infty} (m^p)^{-\tau/2} < \infty,$$

using property (ii) and the fact that $p\tau/2 > 1$. Thus, the Borel–Cantelli Lemma gives

$$(A.43) \quad \xi(\check{\kappa}, \check{\kappa}, n_m) - E\xi(\check{\kappa}, \check{\kappa}, n_m) \rightarrow 0 \text{ a.s. as } m \rightarrow \infty.$$

By definition of $\lambda(\cdot)$ (which relies on Assumption H(ii)),

$$(A.44) \quad \lim_{m \rightarrow \infty} E\xi(\check{\kappa}, \check{\kappa}, n_m) > 0.$$

Equations (A.43) and (A.44) establish property (iv).

Property (v) is established as follows: By (A.39), $\forall n \in [m^p, (m+1)^p]$,

$$(A.45) \quad \begin{aligned} \lambda_{\min} \left[\sum_{i=1}^{m^p} Z_{\check{\kappa}_{m^p}}(X_i) Z_{\check{\kappa}_{m^p}}(X_i)' \right] &\leq \lambda_{\min} \left[\sum_{i=1}^{m^p} Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right] \\ &\leq \lambda_{\min} \left[\sum_{i=1}^n Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right]. \end{aligned}$$

Hence,

$$(A.46) \quad \begin{aligned} &\min_{n \in [m^p, (m+1)^p]} \frac{\lambda(\check{\kappa}_n)}{n} \lambda_{\min} \left[\sum_{i=1}^n Z_{\kappa_n}(X_i) Z_{\kappa_n}(X_i)' \right] \\ &\geq \frac{m^p}{(m+1)^p} \cdot \frac{\lambda(\check{\kappa}_{m^p})}{m^p} \lambda_{\min} \left[\sum_{i=1}^{m^p} Z_{\check{\kappa}_{m^p}}(X_i) Z_{\check{\kappa}_{m^p}}(X_i)' \right] \end{aligned}$$

and property (v) holds. This completes the proof of parts (a) and (b).

To prove part (c) of the Lemma, note that (A.32) and (A.33) both hold with $\lambda_{\min}(\cdot)$ replaced by $\lambda_{\max}(\cdot)$ throughout. Next, redefine $\xi(\kappa_1, \kappa_2, n)$ with $\lambda_{\min}(\cdot)$

and $\lambda(\kappa_{1n})$ replaced by $\lambda_{\max}(\cdot)$ and 1 respectively. Then, properties (i)–(iii) of (A.35) hold as before and properties (iv) and (v) can be replaced by

$$(A.47) \quad \begin{aligned} (iv') \quad & \overline{\lim}_{m \rightarrow \infty} \xi(\check{\kappa}, \check{\kappa}, n_m) < \infty \text{ a.s. and} \\ (v') \quad & \max_{n \in [n_m, n_{m+1}]} \xi(\check{\kappa}, \kappa, n) \leq \xi(\check{\kappa}, \check{\kappa}, n_m) \quad \forall m \geq 1. \end{aligned}$$

Properties (iv') and (v') combine to yield part (c) of the Lemma. Property (iv') is established in the same way as is property (iv) using the fact that $\overline{\lim}_{m \rightarrow \infty} E\xi(\check{\kappa}, \check{\kappa}, n_m) < \infty$ by Assumption I(a). Given (A.39), property (v) is easily seen to hold. \square

The proof of Theorem 3 uses the following lemma concerning Renyi–mixing. This lemma is quite similar to Lemma 5 of Eastwood and Gallant (1987), but it covers inid as well as iid random variables.

LEMMA A-2: Let $\{\epsilon_i : i \geq 1\}$ be a sequence of independent variables. Let $\{w_{ni} : i \leq n, n \geq 1\}$ be a triangular array of non-random weights such that $\max_{i \leq n} |w_{ni}| \rightarrow 0$ as $n \rightarrow \infty$. Let $\{\xi_n\}$ be a sequence of constants such that $\xi_n \rightarrow 0$ as $n \rightarrow \infty$. Define $S_n^* = \sum_{i=1}^n w_{ni} \epsilon_i + \xi_n$. Suppose $S_n^* \xrightarrow{d} F$ as $n \rightarrow \infty$ for some df F . Then,

$$P(S_n^* \leq x | A) \rightarrow F(x) \text{ as } n \rightarrow \infty$$

for all continuity points x of F and for all events A with $P(A) > 0$.

COMMENT: When a sequence of random variables $\{S_n^* : n \geq 1\}$ satisfies the result of Lemma A-2, it is said to be Renyi–mixing, see Renyi (1958).

PROOF OF LEMMA A-2: The proof follows that of Eastwood and Gallant's (1987) Lemma 5. If $F(x)$ equals zero or one, it is trivial to show the result of the Lemma. So assume $F(x) \in (0,1)$ and x is a continuity point of $F(x)$. Let Q_n be the event $(S_n^* \leq x)$. Using Rao (1984, Sec. 8.2, Prop. 2), if (i) $\lim_{n \rightarrow \infty} P(Q_n) = F(x)$, and

(ii) $\lim_{n \rightarrow \infty} P(Q_n | Q_m) = F(x)$ for each m with $P(Q_m) > 0$, then $\lim_{n \rightarrow \infty} P(Q_n | A) = F(x)$ for every event A with $P(A) > 0$.

Condition (i) holds by assumption. To obtain condition (ii), let $S_{nm} = \sum_{i=1}^m w_{ni} \epsilon_i$ for $m = 1, \dots, n$. We have $|S_{nm}| \leq \sup_{i \leq n} |w_{ni}| \sum_{i=1}^m |\epsilon_i| \xrightarrow{p} 0$ as $n \rightarrow \infty$. Hence, $F(x) = \lim_{n \rightarrow \infty} P(Q_n) = \lim_{n \rightarrow \infty} P(S_n^* - S_{nm} \leq x)$ and $\lim_{n \rightarrow \infty} P(Q_n | Q_m)$ $= \lim_{n \rightarrow \infty} P(S_n^* - S_{nm} \leq x | Q_m)$ for every Q_m with $P(Q_m) > 0$. But, since $S_n^* - S_{nm}$ is independent of Q_m , we get $\lim_{n \rightarrow \infty} P(S_n^* - S_{nm} \leq x) = \lim_{n \rightarrow \infty} P(S_n^* - S_{nm} \leq x | Q_m)$ whenever $P(Q_m) > 0$. These results combine to give condition (ii). \square

PROOF OF THEOREM 3: First we establish the asymptotic normality result. Let $T_n^* = A_n^*(\Gamma_n(\hat{g}) - \Gamma_n(g))$, where \hat{g} and A_n^* are as defined in Section 4. Let $T_n^j = A_n(\Gamma_n(\hat{g}) - \Gamma_n(g))$, where \hat{g} and A_n are as defined in (2.5) and (2.9) but with κ_n replaced by κ_n^j . Let B_j be the event $(\{\tilde{\kappa}_n\} = \{\kappa_n^j\})$. Let G_n be the event $(\hat{\kappa}_n = \tilde{\kappa}_n)$. Note that $(T_n^* \leq y) \cap G_n \cap B_j = (T_n^j \leq y) \cap G_n \cap B_j$.

As shown below, Lemma A-2 and Theorem 1(b) imply that $\{c'T_n^j : n \geq 1\}$ is a Renyi-mixing sequence for each $c \in R^V$ and each $j \geq 1$. This yields: $\forall c \in R^V, \forall y \in R$, and $\forall j \geq 1$ such that $P(B_j) > 0$,

$$(A.48) \quad \lim_{n \rightarrow \infty} P(c'T_n^j \leq y | B_j) = \lim_{n \rightarrow \infty} P(c'T_n^j \leq y) = P(c'W \leq y),$$

where $W \sim N(0, I_V)$.

To see that Lemma A-2 applies to $\{c'T_n^j\}$, let Z_j abbreviate $Z_{n\kappa_n^j}$ and write

$$(A.49) \quad c'T_n^j = \sum_{i=1}^n c'A_n \gamma'_{n\kappa_n^j} (Z_j' Z_j)^+ Z_{\kappa_n^j}(x_i) U_i + c'A_n \gamma'_{n\kappa_n^j} (Z_j' Z_j)^+ Z_j' g(X) - c'A_n \Gamma_n(g),$$

$$(A.50) \quad \xi_n = c'A_n \gamma'_{n\kappa_n^j} (Z_j' Z_j)^+ Z_j' g(X) - c'A_n \Gamma_n(g) \rightarrow 0 \text{ as } n \rightarrow \infty$$

by equation (A.5), and

$$\begin{aligned}
 \max_{i \leq n} w_{ni}^2 &= \max_{i \leq n} \left[c' A_n \gamma'_{n\kappa^j} (Z_j' Z_j)^+ Z_{\kappa^j_n} (x_i) \right]^2 \\
 (A.51) \quad &\leq c' A_n \gamma'_{n\kappa^j} (Z_j' Z_j)^+ \gamma_{n\kappa^j} A_n c \cdot \max_{i \leq n} Z_{\kappa^j_n} (x_i)' (Z_j' Z_j)^+ Z_{\kappa^j_n} (x_i) \\
 &\leq (1/\inf_{i \geq 1} \sigma_i^2) \cdot \max_{i \leq n} P Z_{\kappa^j_n} (i) \rightarrow 0 \text{ as } n \rightarrow \infty,
 \end{aligned}$$

by Assumption A and the fact that $\{\kappa_n^j\}$ satisfies Assumption C(ii). Lastly, $c' T_n^j \xrightarrow{d} c' W$ as $n \rightarrow \infty$ by Theorem 1(b).

Next, we have

$$\begin{aligned}
 P(c' T_n^* \leq y) &= \sum_{j=1}^{\infty} P(c' T_n^* \leq y \cap G_n | B_j) P(B_j) + \sum_{j=1}^{\infty} P(c' T_n^* \leq y \cap G_n^c | B_j) P(B_j) \\
 (A.52) \quad &= \sum_{j=1}^{\infty} P(c' T_n^j \leq y \cap G_n | B_j) P(B_j) + P(c' T_n^* \leq y \cap G_n^c) \\
 &= \sum_{j=1}^{\infty} P(c' T_n^j \leq y | B_j) P(B_j) - \delta_n + P(c' T_n^* \leq y \cap G_n^c),
 \end{aligned}$$

where G_n^c denotes the complement of G_n and $\delta_n = \sum_{j=1}^{\infty} P(c' T_n^j \leq y \cap G_n^c | B_j) P(B_j)$.

Using Assumption J we find that

$$(A.53) \quad 0 \leq \delta_n \leq \sum_{j=1}^{\infty} P(G_n^c | B_j) P(B_j) = P(G_n^c) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Thus, taking the limit as $n \rightarrow \infty$ in (A.52) yields

$$\begin{aligned}
 \lim_{n \rightarrow \infty} P(c' T_n^* \leq y) &= \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} P(c' T_n^j \leq y | B_j) P(B_j) \\
 (A.54) \quad &= \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} P(c' T_n^j \leq y | B_j) P(B_j) = P(c' W \leq y),
 \end{aligned}$$

for all $c \in R^V$ and all $y \in R$, where the second equality holds by the bounded convergence theorem, and the third equality holds by (A.48).

Next, we establish the convergence in probability result. By Theorem 1(c), $c'A_n^{-1}T_n^j \xrightarrow{d} 0$ as $n \rightarrow \infty$ for all $j \geq 1$. In addition, the sequence $\{c'A_n^{-1}T_n^j : n \geq 1\}$ is Renyi-mixing for each $c \in R^V$ and each $j \geq 1$, by an argument analogous to that given above. In consequence, equations (A.48), (A.52), (A.53) and (A.54) hold for all $y \neq 0$ with T_n^j , T_n^* , and $c'W$ replaced by $A_n^{-1}T_n^j$, $(A_n^*)^{-1}T_n^*$, and \bar{W} , respectively, where \bar{W} is a random variable with point mass distribution at zero. \square

The following lemma is used in the proofs of Theorems 4 and 5:

LEMMA A-3: Under Assumptions A, D, J, and K(iii),

$$\max_{i \leq n} |\hat{g}(x_i) - g(x_i)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty.$$

PROOF OF LEMMA A-3: Let $\tilde{g}(x)$ denote $\hat{g}(x)$ when $\hat{g}(x)$ is based on $\tilde{\kappa}_n$ terms. Since $P(\tilde{\kappa}_n = \tilde{\kappa}_n) \rightarrow 1$ as $n \rightarrow \infty$ by Assumption J, it suffices to show that $\max_{i \leq n} |\tilde{g}(x_i) - g(x_i)| \xrightarrow{P} 0$ as $n \rightarrow \infty$. By the triangle inequality and Chebyshev's inequality, this holds if

$$(A.55) \quad M_{1n} = \max_{i \leq n} |Z_{\tilde{\kappa}_n}(x_i)'(Z_{n\tilde{\kappa}}'Z_{n\tilde{\kappa}})^+ Z_{n\tilde{\kappa}}'g(X) - g(x_i)| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty \text{ and}$$

$$(A.56) \quad M_{2n} = E \max_{i \leq n} \left[\tilde{g}(x_i) - Z_{\tilde{\kappa}_n}(x_i)'(Z_{n\tilde{\kappa}}'Z_{n\tilde{\kappa}})^+ Z_{n\tilde{\kappa}}'g(X) \right]^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

First we establish (A.55). Conditional on $B_j = (\{\tilde{\kappa}_n\} = \{\kappa_n^j\})$, M_{1n} is non-random and it converges to zero as $n \rightarrow \infty$ by the proof of (A.5) with $b = 1$, $A_n = 1$, $\Gamma_n(g) = g(x_i)$, $E\Gamma_n(\hat{g})$ replaced by $Z_{\kappa_n^j}(x_i)'(Z_{n\kappa^j}'Z_{n\kappa^j})^+ Z_{n\kappa^j}'g(X)$, κ_n replaced by κ_n^j , and $\max_{i \leq n} |\cdot|$ added in the appropriate places. In consequence,

$$(A.57) \quad \begin{aligned} \lim_{n \rightarrow \infty} P(M_{1n} > \epsilon) &= \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} P(M_{1n} > \epsilon | B_j) P(B_j) \\ &= \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} P(M_{1n} > \epsilon | B_j) P(B_j) = 0, \end{aligned}$$

where the second equality holds by the bounded convergence theorem.

Next we establish (A.56). Note that $e_j' Z_{nk} (Z_{nk}' Z_{nk})^+ Z_{nk} e_j$ and $U' Z_{nk} (Z_{nk}' Z_{nk})^+ Z_{nk}' U$ are non-decreasing in k , since each is an explained sum of squared from a projection onto the column space of Z_{nk} . Let \bar{Z} denote $Z_{n\bar{\kappa}}$. We have

$$\begin{aligned}
 M_{2n} &= E \max_{i \leq n} \left[Z_{\bar{\kappa}_n}(x_i)' (Z_{n\bar{\kappa}}' Z_{n\bar{\kappa}})^+ Z_{n\bar{\kappa}}' U \right]^2 \\
 &\leq E \max_{i \leq n} Z_{\bar{\kappa}_n}(x_i)' (Z_{n\bar{\kappa}}' Z_{n\bar{\kappa}})^+ Z_{\bar{\kappa}_n}(x_i) \cdot U' Z_{n\bar{\kappa}} (Z_{n\bar{\kappa}}' Z_{n\bar{\kappa}})^+ Z_{n\bar{\kappa}}' U \\
 (A.58) \quad &\leq E \max_{i \leq n} e_j' \bar{Z} (\bar{Z}' \bar{Z})^+ \bar{Z}' e_j \cdot U' \bar{Z} (\bar{Z}' \bar{Z})^+ \bar{Z}' U \\
 &= \max_{i \leq n} P Z_{n\bar{\kappa}}(i) \cdot \text{tr}(\bar{Z} (\bar{Z}' \bar{Z})^+ \bar{Z}' \Omega) \\
 &\leq \max_{i \leq n} P Z_{n\bar{\kappa}}(i) \cdot \bar{\kappa}_n \sup_{i \geq 1} \sigma_i^2 \rightarrow 0 \text{ as } n \rightarrow \infty,
 \end{aligned}$$

using Assumption K(iii). \square

PROOF OF THEOREM 4: It suffices to show that $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$ as $n \rightarrow \infty$. We have

$$\begin{aligned}
 &\left| \frac{n - \hat{\kappa}_n}{n} \hat{\sigma}^2 - \sigma^2 \right| = \left| \frac{1}{n} \sum_{i=1}^n (U_i + g(x_i) - \hat{g}(x_i))^2 - \sigma^2 \right| \\
 (A.59) \quad &\leq \left| \frac{1}{n} \sum_{i=1}^n (U_i^2 - \sigma^2) \right| + \max_{i \leq n} (\hat{g}(x_i) - g(x_i))^2 + 2 \left[\frac{1}{n} \sum_{i=1}^n U_i^2 \right]^{1/2} \cdot \max_{i \leq n} |\hat{g}(x_i) - g(x_i)| \\
 &= o_p(1) \text{ as } n \rightarrow \infty
 \end{aligned}$$

using Lemma A-3 and the weak law of large numbers, which holds under Assumptions A, K(i), and K(ii). Since $(n - \hat{\kappa}_n)/n \rightarrow 1$ as $n \rightarrow \infty$ under Assumption K(iii), the desired result follows from (A.59). \square

PROOF OF THEOREM 5: Let \hat{V}_{wn}^j and \hat{A}_{wn}^j denote \hat{V}_{wn} and \hat{A}_{wn} when the latter are defined with $\{\hat{\kappa}_n\}$ replaced by $\{\kappa_n^j\}$. Suppose we can show that

$$(A.60) \quad b_n^j \hat{V}_{wn}^j \xrightarrow{p} V_j \text{ as } n \rightarrow \infty$$

for all $j \geq 1$ and $w = 1, \dots, 5$. Then, Assumption L(iii) and standard arguments yield

$$(A.61) \quad \hat{V}_{wn}^j [V_n^j]^{-1} \xrightarrow{P} I_V, \quad [b_n^j]^{-1/2} \hat{A}_{wn}^j \xrightarrow{P} V_j^{-1/2}, \quad \text{and} \quad \hat{A}_{wn}^j [A_n^j]^{-1} \xrightarrow{P} I_V$$

as $n \rightarrow \infty$, for all $j \geq 1$ and $w = 1, \dots, 5$, where $A_n^j = [V_n^j]^{-1/2}$.

Let B_j be the event $(\{\tilde{\kappa}_n\} = \{\kappa_n^j\})$. If $P(B_j) > 0$, then $\lim_{n \rightarrow \infty} P(D_n) = 0$ implies $\lim_{n \rightarrow \infty} P(D_n | B_j) = 0$ for any sequence $\{D_n\}$, since $P(D_n | B_j) \leq P(D_n)/P(B_j)$.

In consequence, (A.61) implies that for all $\epsilon > 0$

$$(A.62) \quad \lim_{n \rightarrow \infty} P \left[\left\| \hat{V}_{wn}^j [V_n^j]^{-1} - I_V \right\| > \epsilon | B_j \right] = \lim_{n \rightarrow \infty} P \left[\left\| \hat{A}_{wn}^j [A_n^j]^{-1} - I_V \right\| > \epsilon | B_j \right] = 0$$

for all j such that $P(B_j) > 0$.

By Assumption J and the bounded convergence theorem,

$$(A.63) \quad \begin{aligned} 1 &= \lim_{n \rightarrow \infty} P(\hat{\kappa}_n = \tilde{\kappa}_n) = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} P(\hat{\kappa}_n = \tilde{\kappa}_n | B_j) P(B_j) \\ &= \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} P(\hat{\kappa}_n = \tilde{\kappa}_n | B_j) P(B_j). \end{aligned}$$

Hence, if $P(B_j) > 0$, then $P(\hat{\kappa}_n = \kappa_n^j | B_j) = P(\hat{\kappa}_n = \tilde{\kappa}_n | B_j) \rightarrow 1$ as $n \rightarrow \infty$ and (A.62) holds with \hat{V}_{wn}^j , V_n^j , \hat{A}_{wn}^j , and A_n^j replaced with \hat{V}_{wn} , V_n^* , \hat{A}_{wn} , and A_n^* respectively. That is, if $P(B_j) > 0$,

$$(A.64) \quad \lim_{n \rightarrow \infty} P \left[\left\| \hat{V}_{wn} (V_n^*)^{-1} - I_V \right\| > \epsilon | B_j \right] = \lim_{n \rightarrow \infty} P \left[\left\| \hat{A}_{wn} (A_n^*)^{-1} - I_V \right\| > \epsilon | B_j \right] = 0.$$

This result and the bounded convergence theorem give

$$(A.65) \quad \begin{aligned} \lim_{n \rightarrow \infty} P \left[\left\| \hat{V}_{wn} (V_n^*)^{-1} - I_V \right\| > \epsilon \right] &= \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} P \left[\left\| \hat{V}_{wn} (V_n^*)^{-1} - I_V \right\| > \epsilon | B_j \right] P(B_j) \\ &= \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} P \left[\left\| \hat{V}_{wn} (V_n^*)^{-1} - I_V \right\| > \epsilon | B_j \right] P(B_j) = 0 \end{aligned}$$

for all $\epsilon > 0$ and $w = 1, \dots, 5$, which establishes the first result of Theorem 5. The second result of Theorem 5 is obtained by replacing \hat{V}_{wn} and V_n^* with \hat{A}_{wn} and A_n^* in

(A.65). The third result of Theorem 5 follows from the second result and Theorem 1(b).

It remains to establish (A.60). For the remainder of this proof all quantities that depend on $\hat{\kappa}_n$, such as $\hat{g}(x_i)$, \hat{U}_i , and $\hat{\Omega}_{wn}$, are taken to be defined with $\hat{\kappa}_n$ replaced by the constant κ_n^j . In addition, Z and γ_j are used to denote $Z_{n\kappa^j}$ and $\gamma_{n\kappa^j}$.

Consider a matrix $\hat{\Omega}_n$ of the form $\text{diag}(\xi_{ni}\hat{U}_i^2)$ for some constants $\{\xi_{ni} : i \leq n, n \geq 1\}$ that satisfy $\max_{i \leq n} |\xi_{ni} - 1| \rightarrow 0$ as $n \rightarrow \infty$. Since $\{\kappa_n^j\}$ satisfies Assumption C(ii) by the definition of \mathcal{K} , the matrices $\hat{\Omega}_{wn}$ are of this form for $w = 1, 2, 3, 5$ and so is the first term of $\hat{\Omega}_{4n}$.

Let $\hat{V}_n^j = \gamma_j'(Z'Z)^+ Z' \hat{\Omega}_n Z(Z'Z)^+ \gamma_j$. For arbitrary $c \in \mathbb{R}^V$ with $\|c\| = 1$, we will show that

$$(A.66) \quad b_n^j c'(\hat{V}_n^j - V_n^j)c = b_n^j \zeta'(\hat{\Omega}_n - \Omega_n)\zeta \xrightarrow{P} 0 \text{ as } n \rightarrow \infty$$

where $\zeta = Z(Z'Z)^+ \gamma_j c \in \mathbb{R}^n$ and

$$(A.67) \quad c' L_n c = c' \left[b_n^j \gamma_j'(Z'Z)^+ Z' \left\{ \frac{n-1}{n^2} \tilde{U} \tilde{U}' \right\} Z(Z'Z)^+ \gamma_j \right] c = o_p(1)$$

as $n \rightarrow \infty$. These results and Assumption L(iii) combine to give (A.60).

It remains to prove (A.66) and (A.67). To show the former, let $a_n = \zeta' \zeta = c' \gamma_j'(Z'Z)^+ \gamma_j c$. Since $\overline{\lim}_{n \rightarrow \infty} b_n^j \lambda_{\max}(V_n^j) < \infty$, we have

$$(A.68) \quad b_n^j \leq C_1^* / \lambda_{\max}(V_n^j) \leq C_1^* / (\inf_{i \geq 1} \sigma_i^2 \cdot \lambda_{\max}(\gamma_j'(Z'Z)^+ \gamma_j)) \leq C_2^* / a_n$$

for some finite constants C_1^* and C_2^* .

With this result, we get

$$\begin{aligned}
& |b_n^j \zeta'(\hat{\Omega}_n - \Omega) \zeta / C_2^*| \leq |\zeta'(\hat{\Omega}_n - \Omega) \zeta / a_n| \\
& = \left| \sum_{i=1}^n \xi_{ni} \zeta_i^2 (\hat{U}_i^2 - \sigma_i^2) / a_n + \sum_{i=1}^n (\xi_{ni} - 1) \zeta_i^2 \sigma_i^2 / a_n \right| \\
(A.69) \quad & \leq \left| \sum_{i=1}^n \xi_{ni} \zeta_i^2 (U_i^2 - \sigma_i^2) / a_n \right| + \sum_{i=1}^n \xi_{ni} \zeta_i^2 (\hat{g}(x_i) - g(x_i))^2 / a_n \\
& \quad + \left| 2 \sum_{i=1}^n \xi_{ni} \zeta_i^2 U_i (\hat{g}(x_i) - g(x_i)) / a_n \right| + \max_{\ell \leq n} |\xi_{n\ell} - 1| \sup_{j \geq 1} \sigma_j^2 \sum_{i=1}^n \zeta_i^2 / a_n \\
& = \left| \sum_{i=1}^n \xi_{ni} \zeta_i^2 (U_i^2 - \sigma_i^2) / a_n \right| + o_p(1) + o_p(1) \sum_{i=1}^n \zeta_i^2 |U_i| / a_n + o_p(1),
\end{aligned}$$

where the second equality uses the fact that $\max_{i \leq n, n \geq 1} |\xi_{ni}| < \infty$, $\sum_{i=1}^n \zeta_i^2 / a_n = 1$, and $\max_{i \leq n} |\hat{g}(x_i) - g(x_i)| = o_p(1)$ by Lemma A-3.

The expression above is $o_p(1)$ provided

$$(A.70) \quad \sum_{i=1}^n \xi_{ni} \zeta_i^2 (U_i^2 - \sigma_i^2) / a_n + o_p(1) \text{ and } \sum_{i=1}^n \zeta_i^2 |U_i| / a_n = O_p(1) \text{ as } n \rightarrow \infty.$$

To establish the two results of (A.70), we need the following inequality:

$$\begin{aligned}
(A.71) \quad & \max_{i \leq n} \zeta_i^2 = \max_{i \leq n} \left[c' \gamma_j' (Z' Z)^+ Z_{\kappa_n^j(x_i)} \right]^2 \\
& \leq a_n \cdot \max_{i \leq n} Z_{\kappa_n^j(x_i)}' (Z' Z)^+ Z_{\kappa_n^j(x_i)} = a_n \cdot o(1).
\end{aligned}$$

Using this inequality and Assumption L(i), we get

$$\begin{aligned}
(A.72) \quad & \text{Var} \left[\sum_{i=1}^n \xi_{ni} \zeta_i^2 (U_i^2 - \sigma_i^2) / a_n \right] = \frac{1}{a_n^2} \sum_{i=1}^n \xi_{ni}^2 \zeta_i^4 E[U_i^2 - \sigma_i^2]^2 \\
& \leq \max_{i \leq n} \xi_{ni}^2 \left(\max_{j \leq n} \zeta_j^2 / a_n \right) \left[\frac{1}{a_n} \sum_{i=1}^n \zeta_i^2 \right] \sup_{j \geq 1} E[U_j^2 - \sigma_j^2]^2 \rightarrow 0 \text{ as } n \rightarrow \infty.
\end{aligned}$$

Since $\sum_{i=1}^n \xi_{ni} \zeta_i^2 (U_i^2 - \sigma_i^2) / a_n$ has mean zero and variance that goes to zero as $n \rightarrow \infty$, it converges in probability to zero, as desired.

Next, we consider the rv $\sum_{i=1}^n \zeta_i^2 |U_i|/a_n$. We have

$$(A.73) \quad \sup_{n \geq 1} E \sum_{i=1}^n \zeta_i^2 |U_i|/a_n \leq \sup_{j \geq 1} E |U_j| \sup_{n \geq 1} \sum_{i=1}^n \zeta_i^2/a_n < \infty \text{ and}$$

$$(A.74) \quad \begin{aligned} \text{Var} \left[\sum_{i=1}^n \zeta_i^2 |U_i|/a_n \right] &= \sum_{i=1}^n \zeta_i^4 \text{Var}(|U_i|)/a_n^2 \\ &\leq (\max_{j \leq n} \zeta_j^2/a_n) (\sup_{j \geq 1} \text{Var}(|U_j|) \sum_{i=1}^n \zeta_i^2/a_n) \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Equations (A.73) and (A.74) yield the second result of (A.70). Thus, (A.66) holds for all c .

We now prove (A.67). Let $H = \text{diag}\{1/(1 - z_{ii}) : i = 1, \dots, n\}$. For any $c \in R^V$ with $\|c\| = 1$, we have

$$(A.75) \quad \begin{aligned} (c' L_n c)^{1/2} &= \left[b_n^j (n-1)/n^2 \right]^{1/2} |c' \gamma_j' (Z' Z)^+ Z' \tilde{U}| \\ &\leq |(b_n^j/n)^{1/2} c' \gamma_j' (Z' Z)^+ Z' H U| + |(b_n^j/n)^{1/2} c' \gamma_j' (Z' Z)^+ Z' H(\hat{g}(X) - g(X))| \\ &= W_{1n} + W_{2n}, \end{aligned}$$

$$(A.76) \quad \begin{aligned} W_{2n} &\leq \left[b_n^j c' \gamma_j' (Z' Z)^+ Z' H^2 Z (Z' Z)^+ \gamma_j c \right]^{1/2} \left[\frac{1}{n} (\hat{g}(X) - g(X))' (\hat{g}(X) - g(X)) \right]^{1/2} \\ &\leq \left[c' (b_n^j V_n^j) c \left[\max_{i \leq n} 1/(1 - z_{ii})^2 \right] / \inf_{i \geq 1} \sigma_i^2 \right]^{1/2} \max_{i \leq n} |\hat{g}(x_i) - g(x_i)| \\ &\xrightarrow{P} 0 \text{ as } n \rightarrow \infty, \end{aligned}$$

and

$$(A.77) \quad \begin{aligned} \text{Var}(W_{1n}) &\leq \frac{b_n^j}{n} c' \gamma_j' (Z' Z)^+ Z \Omega Z (Z' Z)^+ \gamma_j c \cdot \max_{i \leq n} 1/(1 - z_{ii})^2 \\ &= \frac{1}{n} c' (b_n^j V_n^j) c \cdot \max_{i \leq n} 1/(1 - z_{ii})^2 \rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Since $EW_{1n} = 0$, (A.77) implies that $W_{1n} \xrightarrow{P} 0$ as $n \rightarrow \infty$. This result combined with (A.75) and (A.76) yields $c' L_n c \xrightarrow{P} 0$ as $n \rightarrow \infty$, as desired. \square

FOOTNOTES

1. I thank Whitney K. Newey, James L. Powell, Geraldo Souza, two referees, and the co-editor Lars Hansen for very helpful comments and suggestions. I gratefully acknowledge the financial support of the Alfred P. Sloan Foundation and the National Science Foundation through a Research Fellowship and grant numbers SES-8618617 and SES-8821021.
2. Under the assumptions given below, V_n is nonsingular and A_n is well-defined for n large, see the proof of Theorem 1 given in the Appendix.
3. For the case where the multiplicand $(\kappa^{-1}(k)k)^{1/2} \zeta(k)$ is used in Assumption D, the appearance of $\zeta(k)$ implies that $\zeta(k)$ must be finite for all $k \geq 1$.
4. To see this, suppose that conditional on $\{X_i\} = \{x_i\}$ some sequence of random variables $\{\xi_n\}$ converges in distribution to a random variable ξ with probability one, where ξ does not depend on $\{x_i\}$. This implies that for each bounded continuous function f , $E(f(\xi_n)|\{x_i\}) \rightarrow Ef(\xi)$ as $n \rightarrow \infty$. Taking the expectation over $\{X_i\}$ yields the unconditional convergence in distribution of $\{\xi_n\}$ to ξ .
5. Strictly speaking, as used in (2.21), $S_g(R^d)$ denotes the Sobolev smoothness index of the periodic extension of g to R^d rather than the Sobolev smoothness index of g .
6. Cox (1988) actually considers non-random, rather than random, regressors, but requires that they satisfy an asymptotic $U[0,1]$ design. In Remark 3.1, p. 722, he weakens the requirements on the finite sample design to the point where the regressors could be a typical realization of an iid sequence of $U[0,1]$ rv's (provided $p < 1/2$ in his equation (3.5)).
7. As in the previous footnote, Cox actually considers non-random regressors, but ones that behave like typical regressors drawn from an iid sequence of rv's with beta distribution (provided $p < 1/2$ in his equation (1.2)).
8. The results given here for a scalar regressor also apply in the semiparametric regression model with multiple regressors provided the nonparametric part of the regression function depends on a scalar regressor (see Section 5.2 below). The results for a scalar regressor also apply in the additive regression model with multiple regressors (see Section 5.1 below).
9. We prefer to call these models AIR models rather than interaction spline models because the models need not be estimated by splines.

10. If one wishes to estimate the individual functions $\{g_j, g_{jm}\}$ in (5.1), rather than just the regression function g , then identifying restrictions need to be added to these models. For example, one could require that the functions $g_2 \dots, g_d$ integrate to zero over X (with respect to some measure) and that all the functions g_{jm} integrate to zero over X^2 .
11. This is not to be interpreted to be a claim that the dimension d of the regressor vector does not have an effect on the efficiency of estimators in AIR models. Undoubtedly it does. The magnitude of d , however, does not have an effect on the rate of convergence of series estimators in AIR models.
12. This holds provided $F(\cdot)$ is differentiable and the elements of x_i are not functionally related with positive probability.
13. The parameters of these models are identified only up to scale. Hence, "consistency up to scale" is the best one can do with respect to consistency.
14. I thank James MacKinnon for supplying the computer program that was used to carry out these simulations.

REFERENCES

- Abramowitz, M. and I. A. Stegun, eds. (1964): *Handbook of Mathematical Functions*. New York: Dover.
- Andrews, D. W. K. (1989): "Asymptotic Optimality of Generalized C_L , Cross-Validation, and Generalized Cross-Validation for Models with Heteroskedastic Errors," Cowles Foundation Discussion Paper No. 906, Yale University.
- Andrews, D. W. K. and Y. J. Whang (1989): "Additive Interactive Regression Models: Circumventing the Curse of Dimensionality," manuscript in preparation, Cowles Foundation, Yale University.
- Barry, D. (1983): *Nonparametric Bayesian Regression*. Ph.D. Thesis, Department of Statistics, Yale University, New Haven, CT.
- _____ (1986): "Nonparametric Bayesian Regression," *Annals of Statistics*, 14, 934-953.
- Bergstrom, A. R. (1985): "The Estimation of Nonparametric Functions in a Hilbert Space," *Econometric Theory*, 1, 1-26.
- Buja, A., T. Hastie, and R. Tibshirani (1989): "Linear Smoothers and Additive Models," *Annals of Statistics*, 17, 453-510.
- Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975): *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.
- Chamberlain, G. (1986): "Notes on Semiparametric Regression," unpublished manuscript, Department of Economics, Harvard University.
- _____ (1987): "Efficiency Bounds for Semiparametric Regression," unpublished manuscript, Department of Economics, Harvard University.
- Chen, Z. (1988): "Interaction Spline Models and Their Convergence Rates," unpublished manuscript, Department of Statistics, University of Wisconsin, Madison, WI.
- Cox, D. D. (1988): "Approximation of Least Squares Regression on Nested Subspaces," *Annals of Statistics*, 16, 713-732.
- Davidson, R. and J. G. MacKinnon (1985): "Heteroskedasticity-Robust Tests in Regression Directions," *Annales de l'INSEE*, 59/60, 183-218.
- Drygas, H. (1976): "Weak and Strong Consistency of the Least Squares Estimators in Regression Models," *Z. Wahrscheinlichkeitstheorie und verw. Gebiete*, 34, 119-127.
- Eastwood, B. J. (1987): *Confidence Interval Construction in Semi-nonparametric Regression Estimation*. Ph.D. Thesis, Department of Statistics, North Carolina State University, Raleigh, NC.
- _____ (1988): "Asymptotic Normality and Consistency of Semi-nonparametric Regression Estimators Using an Upward F Test Truncation Rule," unpublished manuscript, Statistical Consulting Centre, Carleton University, Ottawa, Canada.

- Eastwood, B. J. and A. R. Gallant (1987): "Adaptive Truncation Rules for Semiparametric Estimators That Achieve Asymptotic Normality," unpublished manuscript, Department of Statistics, North Carolina State University.
- Edmunds, D. E. and V. B. Moscatelli (1977): "Fourier Approximation and Embeddings in Sobolev Space," *Dissertationes Mathematicae*, 145, 1-46.
- Eicker, F. (1963a): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34, 447-456.
- (1963b): "Central Limit Theorems for Families of Sequences of Random Variables," *Annals of Mathematical Statistics*, 34, 439-446.
- (1967): "Limit Theorems for Regressions with Unequal and Dependent Errors," *Proceedings of the Fifth Berkeley Symposium on Probability and Mathematical Statistics*, 1, 59-82.
- Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986): "Semiparametric Estimates of the Relation between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310-320.
- Eubank, R. L. (1988): *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.
- Fabian, V. (1988): "Polynomial Estimation of Regression Functions with the Supremum Norm Error," *Annals of Statistics*, 16, 1345-1368.
- Gallant, A. R. (1981): "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics*, 15, 211-245.
- (1987): "Identification and Consistency in Semiparametric Regression," *Advances in Econometrics*, Fifth World Congress of the Econometric Society, Volume I, ed. by T. F. Bewley. Cambridge, England: Cambridge University Press.
- Geman, S. and C. Hwang (1982): "Nonparametric Maximum Likelihood Estimation by the Method of Sieves," *Annals of Statistics*, 10, 401-414.
- Gu, C., D. M. Bates, Z. Chen, and G. Wahba (1988): "The Computation of GCV Functions through Householder Tridiagonalization with Application to the Fitting of Interaction Spline Models," Technical Report No. 823, Department of Statistics, University of Wisconsin, Madison, WI.
- Härdle, W. and T. M. Stoker (1987): "Investigating Smooth Multiple Regression by the Method of Average Derivatives," Working Paper No. 2004-88, Sloan School of Management, M.I.T., Cambridge, MA.
- Hastie, T. and R. Tibshirani (1986): "Generalized Additive Models" (with comments), *Statistical Science*, 1, 295-318.
- (1987): "Generalized Additive Models: Some Applications," *Journal of the American Statistical Association*, 82, 371-386.

- Hinckley, D. V. (1977): "Jackknifing in Unbalanced Situations," *Technometrics*, 19, 285-292.
- Horn, S. D., R. A. Horn, and D. B. Duncan (1975): "Estimating Heteroskedastic Variances in Linear Models," *Journal of the American Statistical Association*, 70, 380-385.
- Huber, P. J. (1973): "Robust Regression: Asymptotics, Conjectures, and Monte Carlo," *Annals of Statistics*, 1, 799-821.
- Ichimura, H. (1986): "Estimation of Single Index Models," unpublished discussion paper, Department of Economics, M.I.T.
- Kendall, M. and A. Stuart (1979): *The Advanced Theory of Statistics*, Vol. 2, 4th ed. London: Griffin.
- Li, K.-C. (1987): "Asymptotic Optimality for C_P , C_L , Cross-Validation, and Generalized Cross-Validation: Discrete Index Set," *Annals of Statistics*, 15, 958-975.
- MacKinnon, J. G. and H. White (1985): "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305-325.
- Newey, W. K. (1988a): "Adaptive Estimation of Regression Models via Moment Restrictions," *Journal of Econometrics*, 38, 301-339.
- _____. (1988b): "Two Step Series Estimation of Sample Selection Models," unpublished discussion paper, Princeton University, Princeton, NJ.
- Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, forthcoming.
- Powell, M. J. D. (1981): *Approximation Theory and Methods*. Cambridge, England: Cambridge University Press.
- Rao, M. M. (1984): *Probability Theory with Applications*. Orlando, FL: Academic Press.
- Renyi, A. (1958): "On Mixing Sequences of Sets," *Acta Mathematica, Academiae Scientiarum Hungaricae*, 9, 215-228.
- Rice, J. (1986): "Convergence Rates for Partially Splined Models," *Statistics and Probability Letters*, 4, 203-208.
- Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.
- Ruud, P. A. (1986): "Consistent Estimation of Limited Dependent Variable Models Despite Misspecification of Distribution," *Journal of Econometrics*, 32, 157-187.
- Severini, T. A. and W. H. Wong (1987): "Convergence Rates of Maximum Likelihood and Related Estimates in General Parameter Spaces," Technical Report No. 207, Department of Statistics, University of Chicago, Chicago, IL.

- Souza, G. (1987): "On the Asymptotic Normality of Multivariate Least Squares Estimators in the Context of Fourier Flexible Factor Demand Systems," unpublished manuscript, Department of Information and Decision Sciences, University of Illinois at Chicago, Chicago, IL.
- Stein, E. M. (1970): *Singular Integrals and Differentiability Properties of Functions*. Princeton, NJ: Princeton University Press.
- Stock, J. H. (1985): "Nonparametric Policy Analysis," Kennedy School Working Paper, Harvard University.
- Stoker, T. M. (1986): "Consistent Estimation of Scaled Coefficients," *Econometrica*, 54, 1461-1481.
- (1987): "Nonparametric Tests of Additive Derivative Constraints," Discussion Paper No. 2, Measurement and Modelling in Economics Conference, Nuffield College, Oxford.
- Stone, C. J. (1985): "Additive Regression and Other Nonparametric Models," *Annals of Statistics*, 13, 689-705.
- Szegő, G. (1939): *Orthogonal Polynomials*, 1st ed. New York: American Mathematical Society.
- Wahba, G. (1986): "Partial and Interaction Spline Models for the Semiparametric Estimation of Functions of Several Variables," in *Computer Science and Statistics: Proceedings of the 18th Symposium on the Interface*, ed. by T. J. Boardman. Washington, D.C.: American Statistical Association.
- White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838.